Industrial Applications of Lasers

SECOND EDITION

John F. Ready

INDUSTRIAL APPLICATIONS OF LASERS

Second Edition

This Page Intentionally Left Blank

Industrial Applications of Lasers

Second Edition

John F. Ready

HONEYWELL TECHNOLOGY CENTER
MINNEAPOLIS, MINNESOTA

ACADEMIC PRESS
San Diego London Boston
New York Sydney Tokyo Toronto

This book is printed on acid-free paper. ♾

ACADEMIC PRESS
525 B Street, Suite 1900, San Diego, CA 92101-4495, USA
1300 Boylston Street, Chestnut Hill, MA 02167, USA
http://www.apnet.com

ACADEMIC PRESS LIMITED
24-28 Oval Road, London NW1 7 DX, UK
http://www.hbuk.co.uk/ap/

Library of Congress Cataloging-in-Publication Data
Ready, John F., 1932–
Industrial applications of lasers / John F. Ready. — 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-12-583961-8 (alk. paper)
1. Lasers—Industrial applications. I. Title.
TA1677.R4 1997
621.36'6—dc21 96-39078
CIP

ISBN 0-12-583961-8
Printed in the United States of America
97 98 99 00 01 BC 9 8 7 6 5 4 3 2 1

This book is dedicated to my wife Claire.
Without her encouragement,
it would not have been written.

This Page Intentionally Left Blank

Contents

Preface xv
Acknowledgments xvii
Historical Prologue **xix**

Chapter 1 Fundamentals of Lasers **1**

A. Electromagnetic Radiation 1
B. Elementary Optical Principles 4
C. Energy Levels 9
D. Interaction of Radiation and Matter 11
E. Laser Materials 12
F. Population Inversion 15
G. Resonant Cavity 22
Selected References 30

Chapter 2 Properties of Laser Light **31**

A. Linewidth 31
B. Collimation 36
C. Spatial Profiles of Laser Beams 40
D. Temporal Behavior of Laser Output 46
E. Coherence 53
F. Radiance 58
G. Focusing Properties of Laser Radiation 59
H. Power 63
References 63
Selected Additional References 63

Chapter 3 Practical Lasers 66

A. Gas Lasers 66
B. Solid State Lasers 89
C. Semiconductor Lasers 102
D. Organic Dye Lasers 120
References 129
Selected Additional References 129

Chapter 4 Trends in Laser Development 131

A. Semiconductor Lasers 132
B. Diode-Pumped Solid State Lasers 133
C. Chemical Lasers 136
D. Free Electron Lasers 137
E. X-Ray Lasers 139
F. Optical Parametric Oscillators 140
G. Tunsble Lasers 141
References 143
Selected Additional References 143

Chapter 5 Laser Components and Accessories 144

A. Mirrors 144
B. Optics 148
C. Polarizers 149
D. Infrared Materials 151
E. Detectors 152
F. Modulators 165
G. Light Beam Deflectors 171
H. *Q*-Switches 176
I. Nonlinear Optical Elements 179
J. Optical Isolators 182
K. Raman Shifters 183
L. Injection Seeders 184
M. Beam Profilers 185
N. Optical Tables 187
O. Spatial Light Modulators 189
P. Beam Homogenizers 190
Selected References 191

Chapter 6 Care and Maintenance of Lasers 193

A. Damage and Deterioration of Lasers 193
B. Care and Maintenance 208

References 213
Selected Additional References 214

Chapter 7 Laser Safety 215

A. Physiological Effects 216
B. Laser Safety Practices and Standards 224
References 231
Selected Additional References 231

Chapter 8 Alignment, Tooling, and Angle Tracking 232

A. Position-Sensitive Detectors 233
B. Laser Tooling 235
C. Angle Tracking 242
D. Lasers in Construction 244
References 247
Selected Additional References 247

Chapter 9 Principles Used in Measurement 248

A. The Michelson Interferometer 249
B. Beat Production (Heterodyne) 251
C. The Doppler Effect 252
D. Coherence Requirements 253
Selected References 255

Chapter 10 Distance Measurement and Dimensional Control 256

A. Interferometric Distance Measurement 257
B. Laser Doppler Displacement 270
C. Beam Modulation Telemetry 270
D. Pulsed Laser Range Finders 274
E. A Laser Interferometer Application in Mask Production: A Specific Example of Distance Measurement and Dimensional Control 276
References 276
Selected Additional References 277

Chapter 11 Laser Instrumentation and Measurement 278

A. Velocity Measurement 278
B. Angular Rotation Rate 287
C. Diffractive Measurement of Small Dimension: Wire Diameter 295
D. Profile and Surface Position Measurement 297
E. Measurement of Product Dimension 303
F. Measurement of Surface Finish 305
G. Particle Diameter Measurement 307
H. Strain Measurement 309
I. Vibration 310
J. Cylindrical Form Measurement 310
K. Defect Detection 311
L. Surface Flaw Inspection Monitor: A Specific Example 312
M. Summary 313
References 313
Selected Additional References 313

Chapter 12 Interaction of High-Power Laser Radiation with Materials 315

References 334
Selected Additional References 334

Chapter 13 Laser Applications in Material Processing 355

Selected References 342

Chapter 14 Applications of Laser Welding 343

A. Seam Welding: Subkilowatt Levels 345
B. Welding with Multikilowatt Lasers 353
C. Spot Welding 363
D. Specific Examples of Laser Welding Capability 367
E. Summary 369
References 372
Selected Additional References 372

Chapter 15 Applications for Surface Treatment 373

A. Hardening 373
B. Glazing 380
C. Laser Alloying 380

D. Laser Cladding 381
E. Specific Examples of Laser Heat Treating Capability 382
References 383
Selected Additional References 383

Chapter 16 Applications for Material Removal: Drilling, Cutting, Marking 384

A. Laser-Induced Material Removal 384
B. Hole Drilling 387
C. Cutting 395
D. Scribing 409
E. Marking 411
F. Balancing 413
G. Paint Stripping 415
H. Laser Deposition of Thin Films 415
I. Specific Examples of Material Removal 416
References 417
Selected Additional References 417

Chapter 17 Lasers in Electronic Fabrication 419

A. Established Applications in Electronics 419
B. Applications in Integrated Circuit Fabrication 426
C. Summary 432
D. A Specific Example: Laser-Based Photomask Repair 434
References 435
Selected Additional References 436

Chapter 18 Principles of Holography 437

A. Formation of Holograms 437
B. The Holographic Process 442
C. Hologram Types and Efficiency 451
D. Practical Aspects of Holography 456
References 463
Selected Additional References 463

Chapter 19 Applications of Holography 464

A. Holographic Interferometry 464
B. A Miscellany of Applications 480
C. Holographic Optical Elements 486

D. An Example of Holographic Application 487
References 489
Selected Additional References 490

Chapter 20 Laser Applications in Spectroscopy 491

A. Lasers for Spectroscopic Applications 492
B. Types of Laser Spectroscopy 493
C. Applications of Laser Spectroscopy 489
References 509
Selected Additional References 509

Chapter 21 Chemical Applications 510

A. Laser-Initiated Reactions 511
B. Laser-Altered Reactions 512
C. Laser Monitoring of Chemical Dynamics 516
D. Isotope Separation 519
References 529
Selected Additional References 529

Chapter 22 Fiber Optics 530

A. Structures 530
B. Losses 533
C. Manufacture of Optical Fibers 537
D. Connectors, Splicing, and Couplers 538
E. Fiber Amplifiers 540
F. Infrared-Transmitting Fibers 540
G. Fiber Optic Sensors 542
H. Summary 544
References 545
Selected Additional References 545

Chapter 23 Integrated Optics 546

A. Optical Waveguides 547
B. Components for Integrated Optics 551
C. Integrated Optic Circuits 554
D. Applications 557
References 558
Selected Additional References 558

Chapter 24 Information-Related Applications of Lasers 559

A. Lightwave Communications 559
B. Optical Data Storage 566
C. Optical Data Processing 573
D. Laser Graphics 583
E. Consumer Products 585
References 589
Selected Additional References 589

Epilogue A Look at the Future 590

References 593
Selected Additional References 593

This Page Intentionally Left Blank

Preface

This book is intended for people with specific problems to solve. They want to know if lasers can help them. Many books on lasers have provided excellent introductions to the subject of quantum electronics and on specific aspects of laser technology. Fewer have emphasized how lasers have been used for real applications in industry. This book attempts the latter task. It will certainly not solve the reader's problems; rather, it will show how lasers have been used in previous applications. It is hoped that this approach will stimulate the reader, who can then evaluate the potential use of lasers to solve his or her own problem.

This book is not highly mathematical. Equations are given when they can be related to a specific result or when they are useful for understanding the important factors involved in an application, but they are usually simply stated without derivation. The equations in Chapter 19 on holography are an exception. I feel that the more extensive mathematical development there is needed to understand how holograms "really work."

The book defines the jargon that accompanies lasers. It also includes some of the lore that is known and accepted among long-term laser specialists, but that is rarely collected and written down.

To make the book reasonably self-contained, chapters are included about topics relevant to lasers themselves, but without regard to specific applications. These chapters include basic laser fundamentals, unusual properties of laser light, and the types of practical lasers available. Commonly used accessory equipment, needed along with lasers, like modulators, deflectors, and detectors, is described. The important topic of laser safety, necessary for anyone contemplating a laser application, is also included.

The book for the most part emphasizes currently developed applications. Laser welding and cutting, scribing of ceramics, electronic fabrication techniques, and applications in alignment, surveying, and metrology are well established. Lightwave communications based on semiconductor diode lasers has changed the telephone industry substantially. These and other applications are described, with specific

examples given to illustrate the capabilities of the procedures that are described. Laser-based products that have become familiar, like compact disc players and laser printers, are covered.

Future applications in development are also treated. These include items the eventual importance of which is not yet fully established, but that could have big payoffs, like all-optical computers and remote monitoring of the environment. Even farther off, but of great potential importance, is laser-assisted thermonuclear fusion.

A word about units is in order. I have used the units most commonly employed for each application. Thus, in the early chapters on laser fundamentals, the reader will find metric units exclusively. In applications related to production, like welding and drilling, the reader will find substantial use of English units. This has been done purposely, in the face of the growing trend toward metrification. Despite my personal preference for metric units, I believe that English units are easier to use when working with production engineers.

Acknowledgments

I an indebted to the many people who aided in the production of this book. I especially want to thank the personnel of the Communication Resources Department at the Honeywell Technology Center, who worked diligently to prepare many of the line drawings used in the book. I thank all the people and organizations who supplied photographs for use in the book; their identities are given in the captions for the photographs. Finally, I thank all the individuals and organizations who gave permission to reproduce copyrighted material.

This Page Intentionally Left Blank

Historical Prologue

The development of lasers has been an exciting chapter in the history of science and engineering. It has produced a new device with potential for applications in an extraordinary variety of fields. One might begin the history of lasers with Albert Einstein, in 1917. Einstein developed the concept of stimulated emission on theoretical grounds. Stimulated emission is the phenomenon that is utilized in lasers. Stimulated emission produces amplification of light so that buildup of high-intensity light in the laser can occur. The fundamental nature of the stimulated emission process was described theoretically by Einstein.

This characterization of stimulated emission did not lead immediately to the laser. Much additional preliminary work on optical spectroscopy was done in the 1930s. Most of the atomic and molecular energy levels that are used in lasers were studied and investigated during that decade. In retrospect, it seems that by 1940 there was enough information on energy levels and enough development of optical materials to support the invention of the laser. All the necessary concepts had been developed.

During the Second World War, the attention of the technical community was diverted in large measure into the microwave region of the spectrum, and many technical advances involved microwaves. This diversion probably caused the development of the laser to occur in a somewhat roundabout fashion. In 1954, a device called the maser, standing for microwave amplification by stimulated emission of radiation, was developed by Charles Townes and his co-workers. The microwave maser was based on the principle of stimulated emission, and indeed this appears to be the first practical utilization of stimulated emission. There was much enthusiasm and development devoted to masers in the 1950s. However, this work has died out to a large extent, and the main application of the maser today appears to be as a receiver in radioastronomy.

In 1958, Townes and Arthur Schawlow suggested that stimulated emission, which had been utilized in the maser, could also be used in the infrared and visible regions of the spectrum. This suggestion had the effect of returning the attention of

the scientific community to optics, from which it had been diverted for almost two decades. The resulting device was originally termed an optical maser, and much of the early literature on lasers refers to optical masers. That term fairly quickly fell into disuse and was replaced by the simpler term laser, which stands for light amplification by stimulated emission of radiation. This word has endured.

After the suggestion of Townes and Schawlow, a number of laboratories began working on lasers. The first operating laser was the ruby laser, which was developed in 1960 by Theodore Maiman. This first laser emitted a pulsed beam of collimated red light at a wavelength of 694.3 nm. This laser, although primitive by today's standards, immediately captured the attention of the scientific and engineering community. It provided a source of light with properties different from those of conventional light sources. It represented an extremely monochromatic source. The peak power of the pulsed ruby laser was very high. For the first time, the laser offered a source of coherent electromagnetic radiation in the visible portion of the spectrum. It offered many possibilities for applications. Immediately, many more organizations joined in laser research.

The first gas laser, the helium–neon laser, was operated in 1961. As the first continuous laser, this represented another significant advance. The first helium–neon laser operated in the infrared at a wavelength of 1.15 μm. In 1962, the helium–neon laser was operated using a visible transition at 632.8 nm, the first continuous visible laser. For a substantial period, this was the most common type of laser.

Late in 1962, a still different type of laser was invented, the semiconductor diode laser, which first used a small chip of gallium arsenide material. This laser represented a fundamentally different type of laser from the earlier ruby and gas lasers. Today, semiconductor lasers, fabricated from materials like aluminum gallium arsenide and indium gallium arsenide phosphide, are the most common type of lasers, with tens of millions being produced each year.

From 1962 to 1968, much basic development in lasers occurred. Almost all the important types of lasers were invented in this era. In addition, almost all the practical applications of lasers were suggested. The ability of the laser to melt and vaporize small amounts of metal was determined very early. The capability of the laser for welding, cutting, and drilling was recognized. Uses such as communications, data storage, sensing, spectroscopy, interferometry, holography, and all the rest of the multitude of practical applications that are in use today were identified. There was a great deal of enthusiasm for lasers in those early days. The history of the early development of the laser has been well documented [1].

But the status of laser development itself was still rather rudimentary. Lasers that were available were fragile devices, with poor reliability and durability. They were suited for laboratory demonstrations, but not for routine industrial usage. Even though all the applications and their potential benefits had been identified, lasers themselves had not yet developed enough to support the applications well. In some quarters, disillusionment about the laser developed. The period up to about 1968 represents a beginning period of laser development, in which almost all the basic types of lasers were operated, almost all the applications were conceived and

demonstrated, at least in the laboratory, but most of the applications could not be performed satisfactorily in an industrial environment.

At some point in the late 1960s, perhaps around 1968 but extending over a period of a few years, this situation began to change. The engineering development of lasers improved substantially. The existing types of lasers were designed and fabricated with better reliability and durability. By the early 1970s, industrial engineers could purchase long-lived models of lasers and use them in practical applications. It became possible to perform economically, in an industrial environment, many of the applications that had been discussed in the earlier era. By the mid 1970s, lasers had taken their place as a truly practical tool on production lines for cutting, welding, drilling, and marking, in the construction industry for alignment, in the machine tool control business for distance measurement, and for many other types of applications.

During the 1980s and early 1990s, more remarkable developments occurred. Long distance communications was revolutionized by the use of semiconductor lasers and fiber optics in lightwave communications. The laser became familiar in a variety of consumer products, like compact optical disks, laser printers, and barcode scanners. The development of laser technology for separation of uranium isotopes has been transferred from government laboratories to private industry. By the mid-1990s, the worldwide commercial sales of lasers exceeded 1.4 billion dollars [2]. This represents only the lasers; it does not include all the related accessory equipment that accompanies them.

Still, some suggested applications remain in a developmental stage. Laser-assisted thermonuclear fusion is still advancing, but it is not near practical power generating status. Remote laser sensing of the environment is being utilized in some areas, but it has not reached its full potential. Digital optical computing is still developing. There are exciting future prospects for x-ray lasers. We may confidently expect laser technology to continue to grow and develop, both in the capabilities of the lasers themselves and in the applications they perform. There remain many exciting and important challenges for the future in laser technology.

References

[1] J. L. Bromberg, *The Laser in America, 1950–1970*. MIT Press, Cambridge, MA, 1991.

[2] S. G. Anderson, *Laser Focus World*, p. 50 (January 1996).

This Page Intentionally Left Blank

Chapter 1 | Fundamentals of Lasers

This chapter will provide fundamental information about lasers. It will describe the basic principles of laser operation and will define some of the terminology commonly encountered in laser work. The level is elementary; the chapter is intended as an introduction for those who have not worked extensively with lasers.

A. Electromagnetic Radiation

The light emitted by a laser is electromagnetic radiation. We note here that the word "light" will be used in an extended sense, including infrared and ultraviolet as well as visible light. Also, the word "radiation" refers to radiant energy and does not imply ionizing radiation.

The term "electromagnetic radiation" includes a continuous range of many different types of radiant energy. This section will describe the spectrum of electromagnetic radiation, and it will show how laser light is related to other familiar forms of radiant energy.

Electromagnetic radiation has a wave nature. The waves consist of oscillating electric and magnetic fields. The waves can be characterized by their frequency, that is, by the number of cycles of oscillation of the electric or magnetic field per unit time. The distance between the peak amplitudes of the oscillating field is called the wavelength. The wave propagates with a characteristic velocity c ("the velocity of light") equal to 3×10^{10} cm/sec in vacuum. The physical nature of the electromagnetic radiation is the same in all portions of the electromagnetic spectrum. It has the same wave nature and the same velocity, c. It differs only in wavelength and frequency. The different regions in the electromagnetic spectrum are characterized by different values of wavelength and frequency for the oscillation of the wave.

There is a relation between the frequency f and wavelength λ valid for all types of electromagnetic radiation:

$$\lambda f = c \tag{1.1}$$

According to this relation, the wavelength decreases as frequency increases.

Figure 1-1 shows the electromagnetic spectrum, with several familiar regions identified, along with a wavelength and a frequency scale. Radio waves are at the low-frequency, long-wavelength end of the spectrum. As frequency increases and wavelength decreases, we pass through the microwave, the infrared, the visible, the ultraviolet, the x-ray, and the gamma ray regions. The frequency varies by a factor of 10^{16} or more. The boundaries between the different regions are not sharp and distinct; rather, they are defined by convention according to how the radiation interacts with materials. We emphasize that all these types of radiation are essentially the same in their nature, differing in the frequency of the oscillation of the wave motion.

In addition to its wave characteristics, electromagnetic radiation also has a particle-like nature. In some cases, light acts as if it consisted of discrete particle-like quanta of energy, called photons. Each photon carries a discrete amount of energy. The energy E of a photon is

$$E = hf = \frac{hc}{\lambda} \tag{1.2}$$

where h is Planck's constant, equal to 6.6×10^{-27} erg-sec. Equation (1.2) indicates that the photon energy associated with a light wave increases as the wavelength decreases.

In many of the interactions of light with matter, the quantum nature of light overshadows the wave nature. The light can interact only when the photon energy exceeds some characteristic value. In such interactions, it is difficult to reconcile conceptually the wave and particle natures of light. In some experiments, for example, those involving diffraction and interference, the wave nature will dominate. In other experiments, like those involving absorption of light by atomic and molecular systems, the photon nature will dominate. Thus, light will be absorbed between energy levels of a molecule only when the photon energy equals the energy difference between the levels.

A classic example of the photon nature of light involves the photoelectric effect, in which electrons are emitted from a surface that absorbs light. There will be a minimum photon energy required for emission of an electron. This value is called the work function Φ of the surface and is a function of the material. If the photon

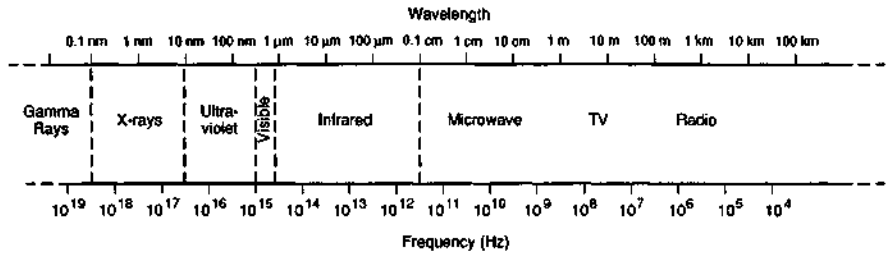


Figure 1-1 The electromagnetic spectrum, with wavelength and frequency indicated for different spectral regions.

energy slightly exceeds this value, that is, if the wavelength is slightly shorter than hc/Φ, electrons are emitted. But if the wavelength is slightly longer than this value, no photoelectric emission occurs, even if the light is intense.

The fact that light has both a wave and a particulate character is called the dual nature of light. This phenomenon can be conceptually troublesome. For our purposes, we simply point out that light sometimes behaves like a wave and at other times like a particle.

With lasers, we are concerned with light in the ultraviolet, visible, and infrared portions of the spectrum, that is, with wavelengths in the approximate range 10^{-5}–10^{-2} cm and frequencies of the order of 10^{13}–10^{15} Hz. The wavelength of light is usually expressed in units of micrometers (μm) or nanometers (nm). The micrometer, equal to 10^{-4} cm, is sometimes called a micron, especially in older literature. The angstrom (Å), a unit equal to 10^{-8} cm, is also sometimes encountered. The wavelength of green light may be given as 0.55 μm = 550 nm = 5.5×10^{-5} cm = 5500 Å.

The complete range of wavelengths covered by operating lasers runs approximately from 0.01 to 1000 μm, but at the far ends of this range, the existing devices are experimental laboratory systems. The region in which useful devices for industrial applications operate is about 0.2 to 10 μm. This region is shown in Figure 1-2, which shows an expanded view of part of the electromagnetic spectrum. The wavelengths of several popular lasers are noted. The wavelength of a given laser is sharply defined; the spread in wavelength for a particular laser covers a very small fraction of the wavelength range shown in the figure.

Lasers can be fabricated using a variety of different materials as the active medium. Each such material has its own distinctive wavelength(s). Figure 1-2 introduces some of these materials and shows where they occur in relation to each other

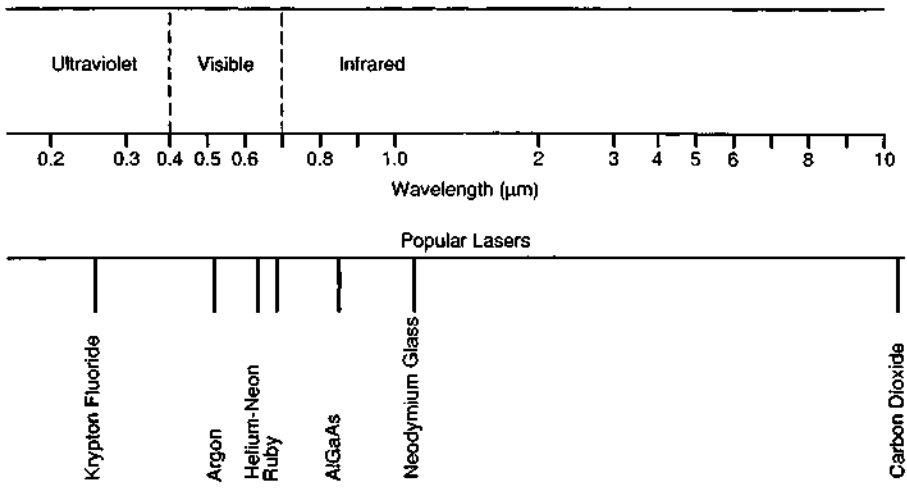


Figure 1-2 Expanded portion of the electromagnetic spectrum, showing the wavelengths at which several important lasers operate.

and to the electromagnetic spectrum as a whole. These useful lasers include the following:

- The CO_2 laser, operating at 10.6 μm in the far infrared
- The neodymium glass laser in the near infrared at 1.06 μm
- The aluminum gallium arsenide laser in the near infrared at wavelengths around 0.85–0.90 μm
- The ruby laser, in the deep red part of the visible spectrum, at 0.6943 μm
- The helium–neon laser, emitting reddish-orange light at 0.6328 μm
- The argon laser, operating at several wavelengths in the blue and green portions of the spectrum
- The krypton fluoride laser, an ultraviolet laser, operating near 0.249 μm

This enumeration does not exhaust the list of important lasers, but it does introduce some significant ones.

B. Elementary Optical Principles

In this section we shall briefly summarize several important optical phenomena that are commonly encountered in laser applications. This is not a comprehensive course in optics; rather, it is a brief review of some optical principles that we will use later. These phenomena will be familiar to students of optics. The three phenomena that we shall emphasize will be the polarization, diffraction, and interference of light beams. These phenomena are all based on the wave nature of light. Phenomena such as reflection, refraction, and index of refraction will be assumed to be familiar. The reader may refer to any elementary optics text.

An instantaneous snapshot of the wave amplitude of a linearly polarized beam of light is sketched in Figure 1-3. The electromagnetic wave is characterized by an electric field and a magnetic field, oriented perpendicular to each other and both perpendicular to the direction of propagation. The pattern sketched in Figure 1-3 will travel along the direction of propagation at the velocity c. At a particular point in the beam of light, the electric field vector **E** will oscillate along a given line in a plane perpendicular to the direction of propagation. This situation is referred to as linear polarization, or plane polarization. An observer at a fixed point will see an electric field oscillating with its orientation in this specific direction. Unpolarized light, on the other hand, will have many components of electric field vectors, oriented in all directions perpendicular to the direction of propagation. In this case, the instantaneous electric field seen by a fixed observer will move randomly in the plane perpendicular to the direction of propagation.

If one combines two beams with linear polarization at right angles to each other, and with constant phase difference, one has elliptically polarized light. The fixed observer will see the tip of the electric field vector moving around an elliptical path in the plane. The ellipse reduces to a circle if the amplitudes of the two beams are equal and the phase difference is $\pi/2$ (or an odd integral multiple of $\pi/2$). The light is then said to be circularly polarized. There are two senses to the circular polarization:

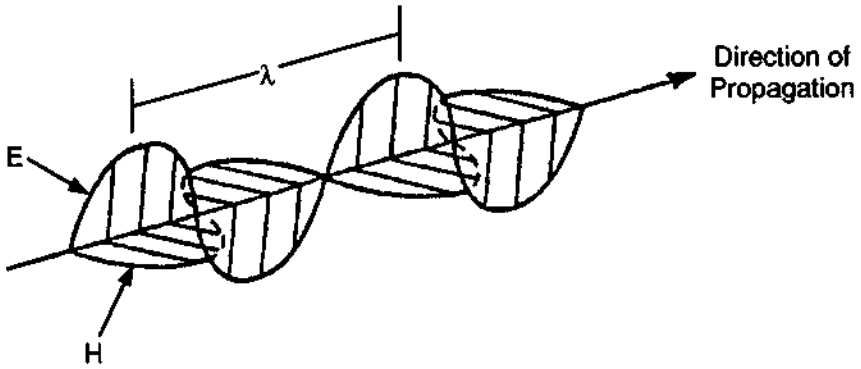


Figure 1-3 Wave amplitude of a linearly polarized beam of light at one instant of time, showing the wavelength λ and the components of electric field vector **E** and magnetic field vector **H**.

right-handed if the direction of rotation of the tip of the electric field vector is clockwise when one looks opposite to the direction in which the light travels, or left-handed if the rotation is counterclockwise.

Laser beams are often linearly polarized. The polarization of a beam can be tested by inserting a polarizer in the beam path. Polarizers can take many forms, some of which are described in Chapter 5. The most familiar is the common polaroid sheet. A polarizer has an oriented axis, called the transmission axis. It transmits light with electric field vector parallel to this axis. If a polarizer is inserted in a beam of linearly polarized light, maximum transmission will occur when the transmission axis is aligned with the direction of the electric field vector. When the polarizer is rotated, the transmitted light intensity will vary as the square of the cosine of the angle between the transmission axis and the electric field vector. If the beam is unpolarized, there will be no variation in intensity of the transmitted light as the polarizer is rotated.

Polarization effects are often used to eliminate losses when light passes from vacuum or a gas into a transparent solid. At such a boundary, when light is incident perpendicular to the solid surface, a fraction of the light equal to $(n-1)^2 / (n+1)^2$ will be reflected. Here, n is the index of refraction of the solid. If n is equal to 1.5, characteristic of glass, 4 percent of the light will be reflected at the interface. This reflection usually represents an undesirable loss for a laser system. The loss can be eliminated (or at least minimized) by proper orientation of the interface and the direction of polarization. In Figure 1-4, light is incident on an interface, assumed to be air–glass for specificity, at an angle θ to the normal to the surface. The direction of the light beam and the normal to the surface define a plane P. Two components of the polarization vector may be considered, the so-called s-component, which is perpendicular to plane P, and the p-component, which lies in plane P. Figure 1-4a shows this orientation.

The reflectance at the interface is shown in Figure 1-4b as a function of θ. At normal incidence, $\theta = 0°$, the loss is the same for each component and is equal to 4 percent. The loss of the s-component increases with θ, approaching 100 percent at

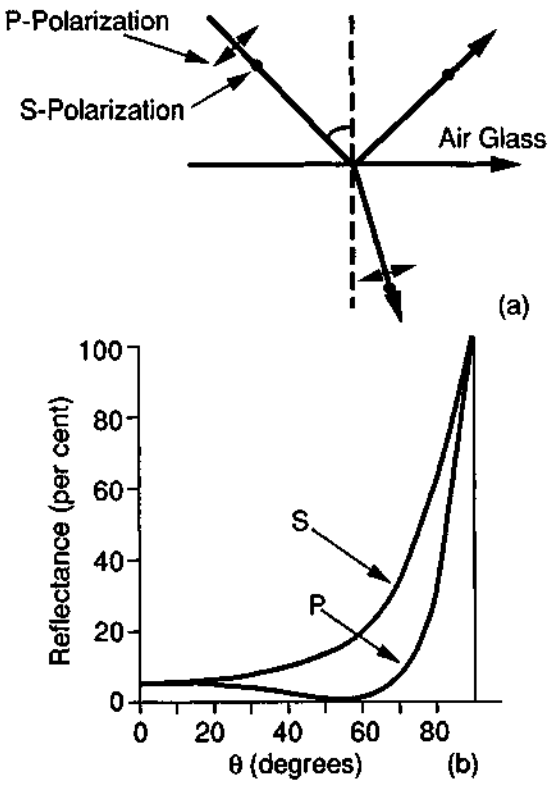


Figure 1-4 Reflectance as a function of incidence angle θ for a light beam traveling from air into glass with index of refraction 1.5. Part (a) defines the geometrical orientation and the p- and s-components of polarization. Part (b) shows the reflectance for each of these components as a function of the angle of incidence, indicating a minimum at Brewster's angle for the p-component.

grazing angles near 90°. The p-component has a minimum of zero at an angle θ_B, called Brewster's angle. Brewster's angle is given by

$$\tan \theta_B = n \tag{1.3}$$

where n is the index of refraction. For ordinary glass, $n \approx 1.5$ and $\theta_B \approx 56°$. Thus, proper orientation of the direction of polarization and the surface normal can eliminate the reflection loss when light enters a transparent solid.

Birefringence is another important phenomenon related to polarization. In certain transparent crystalline materials, the velocity of a light wave depends on its polarization. Such crystals are said to be birefringent, or doubly refracting. This is illustrated in Figure 1-5. When the beam of unpolarized light enters the birefringent crystal, the two different components of polarization propagate in two slightly different directions, as indicated. The difference arises because of the difference in velocity for the two components of polarization. There will be two separate beams emerging from the crystal, one polarized horizontally and one polarized vertically.

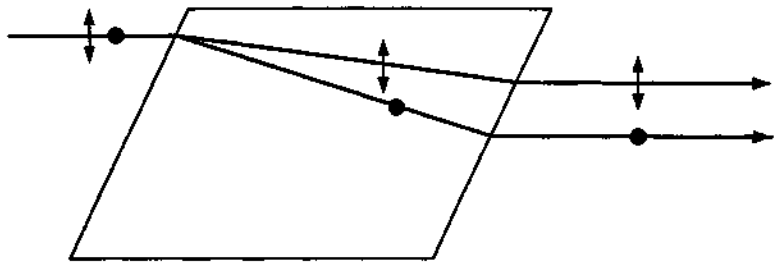


Figure 1-5 Ray directions for two components of polarization in an unpolarized beam of light entering a birefringent crystal.

Birefringence may be demonstrated easily if one places a piece of birefringent crystal, like calcite, on a sheet of printed paper. Two different images of the print will be seen through the crystal. Each image is produced by light of a different polarization. If one places a polarizer above the crystal and rotates it, the difference in polarization will be readily apparent.

In some materials, birefringence may be induced by application of voltage to the crystal. There are a number of such crystalline materials, for example, potassium dihydrogen phosphate, commonly called KDP. This phenomenon is called the electrooptic effect.

The ability of birefringence to separate two perpendicular components of polarization and have them travel in different paths leads to many applications in devices associated with lasers. The electrooptic effect makes possible many useful control functions for laser light. Thus, birefringence, either natural or electrically induced, is commonly used in polarizers, modulators, switches, and light beam deflectors or scanners. These devices are often used in support of laser applications. Chapter 5 offers a more complete description of the use of birefringence in these devices.

A second important phenomenon is diffraction, which is the bending of light around the edges of an obstacle in the beam path, so that light enters the region that would be considered to be the geometrical shadow. As an example, we consider the shadow of a straight edge. According to geometrical optics, if a straight edge is placed between a light source and a screen, the edge would produce a sharp shadow on the screen. No light would reach the points within the shadow, and outside the shadow, the screen would be uniformly illuminated. However, careful observation of the shadow shows that the boundary region is characterized by alternating bright and dark bands, and that a small amount of light has been bent or diffracted around the edge so that light enters the shadow.

Diffraction effects have long been known, but they become more important for a collimated laser beam. In many cases with conventional light sources, the beam divergence is large enough to obscure the fringes formed by diffraction. The collimated nature of laser light makes diffraction effects much easier to observe. This has the twofold effect of providing practical applications based on diffraction, and the sometimes troublesome necessity of minimizing diffractive effects where they are not desired.

Perhaps the most important case of diffraction for lasers involves diffraction by a circular aperture. This is important because the exit aperture of the laser, often circular, defines an obstacle that will diffract the emerging beam. The practical implications of this will be considered in greater detail in Chapter 2. The angular spread of the light emerging from a uniformly illuminated circular aperture of radius a is proportional to $[J_1(ka \sin \theta)/(ka \sin \theta)]^2$, where θ is the angle between the normal to the aperture and the direction of observation, $k = 2\pi/\lambda$, and $J_1(x)$ denotes a Bessel function of order one. The behavior of $[J_1(x)/x]^2$ is similar to $[(\sin x)/x]^2$. This is an alternating pattern of bright and dark rings. The light is most intense in the center of the pattern, $\theta = 0$, and falls to zero at the first dark ring, where $ka \sin \theta$ equals the first zero of $J_1(x)$. Succeeding smaller oscillations provide alternating bright and dark rings surrounding the central pattern. One may define the beam divergence as the angle corresponding to the first dark ring. The first zero of the Bessel function $J_1(x)$ occurs at $1.22\pi = 3.83$. This means that the angle of divergence θ_d will be given by

$$\theta_d \approx \frac{1.22\lambda}{2a} \tag{1.4}$$

As the aperture size decreases, the angular spread of the beam increases.

The third phenomenon of interest is interference. When two or more trains of light waves intersect each other, they interfere so that the electric field at a point is a superposition of the fields in the different waves. The resultant electric field at any point and time is the sum of the instantaneous fields from the individual waves. If the summation produces a zero or small value of the field, the waves are said to interfere destructively at that position, and the light intensity there is low. If the electric fields add to a relatively large value, they are said to interfere constructively, and the light intensity at that point will be relatively high. Thus, interference effects can produce regions of high and low light intensity.

As an example, consider interference produced by light reflected from a thin wedged film, as illustrated in Figure 1-6. A ray of light is incident on the upper surface of the film. Part of the light is reflected at the upper surface, and part is transmitted, The transmitted light simultaneously undergoes refraction, so that it is bent toward the normal to the surface. At the lower surface, part of the light is again reflected, and a fraction of this light will emerge from the top of the film in an upward direction. This light interferes with the light originally reflected from the top surface. At points where the total round-trip optical path difference for these two rays is equal to an odd integral number of half-wavelengths, the electric field vectors will be out of phase with each other and will add to zero. The condition for this is

$$2nd = (2N+1)\frac{\lambda}{2} \tag{1.5}$$

where d is the physical thickness of the film, n is the index of refraction, and N is an integer. (We note that the optical thickness of the film is nd.)

At positions where the wedge is one-quarter wavelength thicker, the two reflected waves will be in phase with each other. At these positions,

$$2nd = N\lambda \tag{1.6}$$

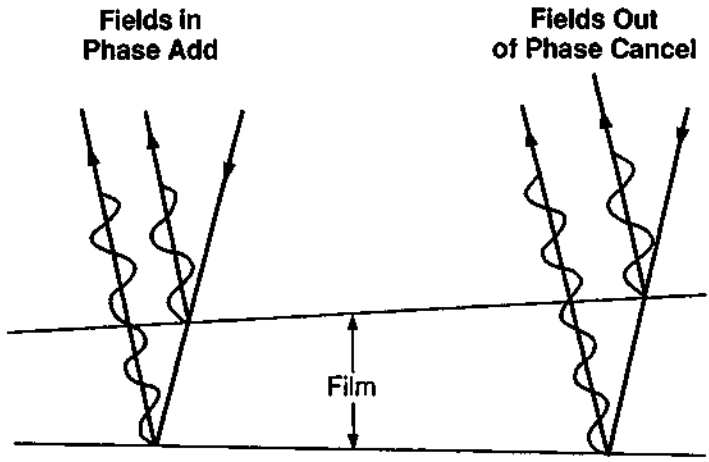


Figure 1-6 Schematic diagram of interference produced in reflection of light from a thin wedged film, showing constructive and destructive interference at positions of different thickness in the film.

where N is again an integer; that is the round-trip optical path difference is an even integral number of half-wavelengths. The electric fields will be in phase and will add to produce maximum light intensity. Thus, the pattern will be a system of bright and dark fringes, with the period of the fringes being the distance along the plate in which the thickness increases by a quarter-wavelength.

Interference fringes can be observed under many other sets of conditions. They have been known for many years, but again the availability of the laser has made the observation and application of interference phenomena much easier. There are many practical applications involving the use of interference effects.

C. Energy Levels

The concept of an energy level is important for lasers. Laser operation takes place via transitions between different energy levels of an atomic or molecular system.

The simplest types of energy levels are those available to an isolated atom. The hydrogen atom is a familiar example. The electron in a hydrogen atom can take up certain discrete orbits around the nucleus. From a classical point of view, one might expect that a continuum of orbits would be available. But quantum mechanics shows that only certain discrete, quantized orbits are possible. In the hydrogen atom, each of these orbits corresponds to a specific energy level. The electron in a given orbit possesses a definite value of energy. The different values of energy that the electron can have, corresponding to the different quantized orbits, are the energy levels of the hydrogen atom. Similarly, other isolated atoms all possess their own characteristic sets of energy levels.

Figure 1-7 shows a partial diagram of the energy levels of the hydrogen atom. The energy scale can be expressed in units of energy (electron volts) or in units of reciprocal length (also called wave numbers), because as Equation (1.2) shows, energy is inversely proportional to wavelength.

When one considers a material such as a solid, liquid, or gas, the energy levels are no longer those of the individual atoms or molecules. An atom or molecule interacts with its neighbors, and the energy levels of the individual atom or molecule are modified by the interactions.

In a gas with relatively low density, the interaction is not strong, and the energy levels are similar to those of the isolated atom or molecule. In a condensed material, solid or liquid, the atoms are closer together, and the interaction becomes strong. The energy levels broaden and merge into a nearly continuous band, consisting of a

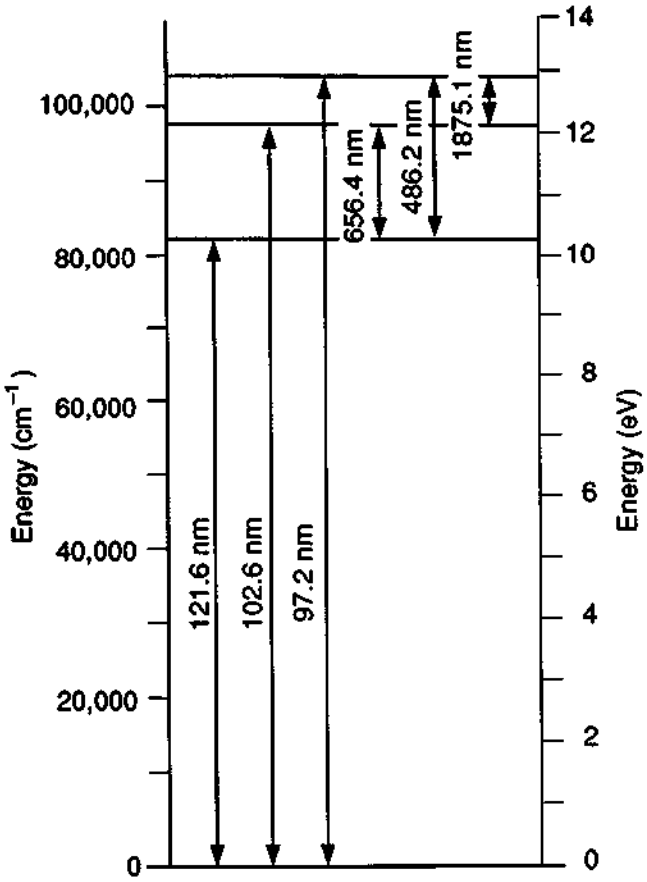


Figure 1-7 Partial energy level diagram of the hydrogen atom, with wavelengths indicated for several transitions. Energy is indicated both in energy units (electron volts) and in wavenumber units (reciprocal centimeters).

large number of closely spaced energy levels. This is, for example, the origin of the familiar conduction and valence bands in semiconductors.

One important exception to the broadening of energy levels in solids is the case where energy levels belong to an unfilled shell of electrons, surrounded by an outer shell of electrons. Examples are the transition metal elements, like titanium, vanadium, chromium, and manganese, and the rare earth elements, like neodymium, erbium, and holmium. These electronic energy levels are shielded from the interaction of the surrounding atoms by the outer shell of electronic charge; the interaction does not broaden the energy levels as much as is usual in solids. The narrow energy levels in certain rare earth elements and transition metals are important for laser operation.

In addition to electronic energy levels, molecules with more than one atom can possess additional energy levels associated with vibration or rotation of the molecule. An example is the carbon dioxide molecule, a linear arrangement of two oxygen atoms and a carbon atom, with the carbon in the middle. Vibrational energy levels correspond to motion of the oxygen atoms relative to the carbon atom. Such energy levels are important for the operation of many gaseous laser systems. Rotational energy levels correspond to rotational motion of an asymmetric molecule. Typically, the ordering of the energy is such that electronic energies are largest, vibrational energies are second, and rotational energies have the smallest magnitude. Thus, lasers based on electronic energy levels usually operate in the ultraviolet, visible, or near infrared portions of the spectrum, and lasers based on vibrational levels operate in the mid or far infrared parts of the spectrum. Rotational energy levels often manifest themselves by causing additional structure in the lines from a laser based on vibrational levels.

D. Interaction of Radiation and Matter

When radiation (light) interacts with matter, it involves a change from one energy level to another. The energy difference between the levels must be balanced by the emission or absorption of radiant energy. The relevant relation is

$$hf = \frac{hc}{\lambda} \approx E_2 - E_1 \tag{1.7}$$

where E_2 and E_1 are the energies of the two states involved (numbered so that $E_2 > E_1$), and hf is the photon energy.

We may distinguish three different ways in which radiation (light) can interact with energy levels: fluorescence, absorption, and stimulated emission. These phenomena demonstrate the particulate nature of light. Whenever light is absorbed or emitted by an atomic or molecular system, it always involves a quantum of energy as specified by Equation (1.7).

1. FLUORESCENCE

An atom in an upper energy level spontaneously decays to a lower energy level, emitting the energy difference as a photon with the appropriate frequency to satisfy

Equation (1.7). As an example, an electron in a hydrogen atom may shift from one orbit to another with a lower energy. The difference in the energies of the two orbits is emitted as a photon with the proper frequency to satisfy Equation (1.7). The wavelengths corresponding to several transitions in the hydrogen atom have been shown in Figure 1-7.

2. *ABSORPTION*

Light of frequency f interacts with an atom or molecule in a low-lying energy level and raises it to a higher level. The light energy is absorbed by the atom or molecule. Unless the frequency of the light satisfies Equation (1.7) for some pair of energy levels, there will be no absorption, and the material will be transparent at that frequency. Because many materials have numerous closely spaced energy levels, a suitable combination of energy levels can exist for almost any wavelength and the material will be opaque. Absorption of light by materials is a familiar phenomenon.

3. *STIMULATED EMISSION*

An atom in an upper energy level interacts with incoming light of frequency f satisfying Equation (1.7), and the atom is stimulated to drop to a lower energy level and to emit the energy difference as light. This process is much less familiar than either fluorescence or absorption, but it is the process responsible for laser operation. It was first postulated by Albert Einstein, who showed that emission of light could be triggered by an incoming photon if an electron were present in a state with high energy. The photon must have the proper energy corresponding to the energy difference between the original state and a lower energy level.

The result of the stimulated emission process is that one photon interacts with an atomic system and two photons emerge. Both photons have the same frequency, the same direction of travel, and the same phase for their associated electric fields. Energy has been extracted from the atom and appears as additional light. The original light is still present, and so the light intensity has been amplified. This is the origin of the acronym LASER: Light Amplification by Stimulated Emission of Radiation.

These three processes are shown schematically in Figure 1-8. The horizontal straight lines represent energy levels; the wavy arrows represent photons; and the vertical arrows represent the transition of the electron from one level to another.

We have now identified the basic stimulated emission process responsible for laser operation. To use this process, one requires the following things: some material with an appropriate energy level structure, a method for exciting the material so as to have electrons present in high-lying energy levels, and a structure called a resonant cavity. In the next sections we shall describe each of these requirements in turn.

E. Laser Materials

A laser needs a material with a suitable set of energy levels. The material is often called the active medium for the laser. The active medium can be a solid, liquid, or

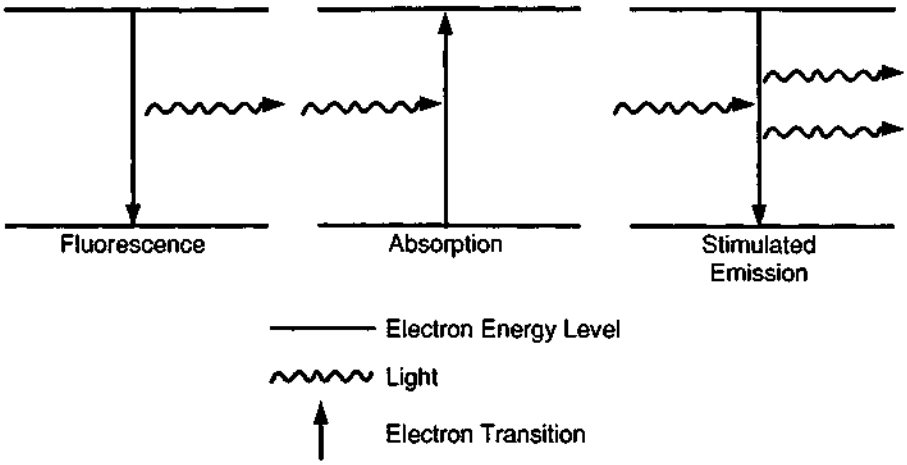


Figure 1-8 Schematic diagram of the interaction of light with electronic energy levels in the processes of fluorescence, absorption, and stimulated emission.

gas, but it must have a suitable set of energy levels in order to interact with light via the stimulated emission process. Laser operation occurs at wavelengths at which the material has fluorescent emission, that is, at frequencies satisfying Equation (1.7) for some pair of energy levels.

Examples of materials with sets of energy levels that have proved useful include ruby (aluminum oxide crystal doped with a small amount of chromium), glass doped with a small percentage of the rare earth element neodymium (Nd:glass), a type of garnet crystal doped with neodymium (referred to as Nd:YAG), carbon dioxide gas, neon gas, ionized argon gas, and the semiconductor crystal aluminum gallium arsenide.

Potentially useful laser materials are studied spectroscopically to determine their energy levels. Spectroscopically, one desires materials with strong fluorescent emission, and with narrow spectral width for the fluorescence. Some specific examples of energy levels are displayed in Figure 1-9 for a solid material and in Figure 1-10 for a gas. Figure 1-9 shows the important energy levels for ruby, with its laser transition at 0.6943 μm. Figure 1-10 shows the relevant energy levels of neon, with three laser transitions identified. The levels are labeled with their spectroscopic designations (e.g., $^3s_2$), which need not concern us here. The energy scale is expressed in electron volts (eV) relative to the lowest (or ground) level as zero. We note that 1 eV is equal to 1.6×10^{-12} erg.

Figures 1-9 and 1-10 also contain additional information about the way is which the excitation occurs. We shall return to this later.

Some discussion of commonly encountered jargon follows. The fluorescent line that is used for laser operation has a finite spectral width. The distribution of

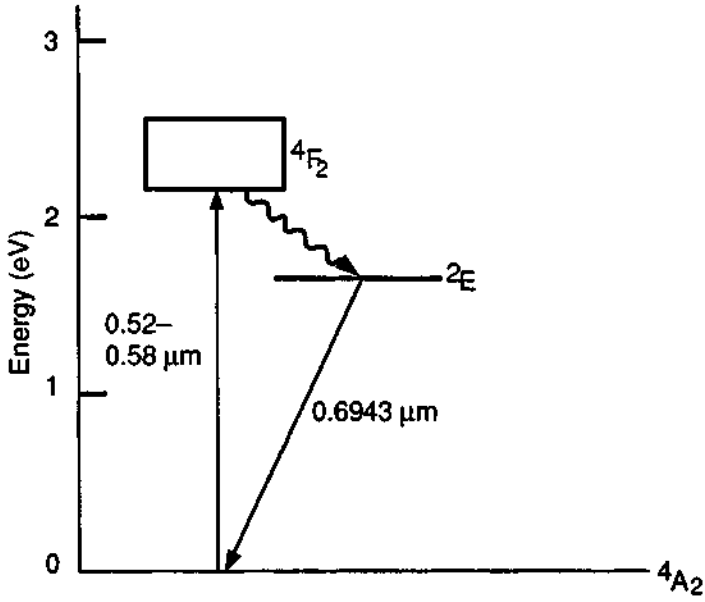


Figure 1-9 Partial energy level diagram of ruby, showing the spectroscopic notation for the relevant energy levels for absorption of green light (0.52–0.58 μm) and spontaneous emission of red light (0.6943 μm).

frequencies in the line defines the fluorescent line shape. There are two common line shapes: Lorentzian and Gaussian. The Lorentzian distribution function, which arises from the finite fluorescent lifetime of the emitting atoms, varies as

$$[(f-f_0)^2+(\Delta f)^2]^{-1}$$

where f is the frequency, f_0 the center frequency of the fluorescent line, and Δf the half-width of the line. The Gaussian distribution function, which arises from Doppler shifts of moving molecules in a gas, varies as

$$\exp\left\{-(\ln 2)\left[\frac{(f-f_0)}{\Delta f}\right]^2\right\}$$

These distribution functions define the intensity of a fluorescent line as a function of frequency.

There is another distinction between line shapes: homogeneous or inhomogeneous. In a homogeneous line, any atom can emit over the whole width of the line; in an inhomogeneous line, any one atom can emit only in a small region within the spectral line width. Thus, for laser operation at some particular frequency within the fluorescent line, the entire population of atoms can contribute in a homogeneous line, but only a small fraction of the atoms can contribute in an inhomogeneous line. When this fraction of the atoms has contributed its energy, one says that "a hole has

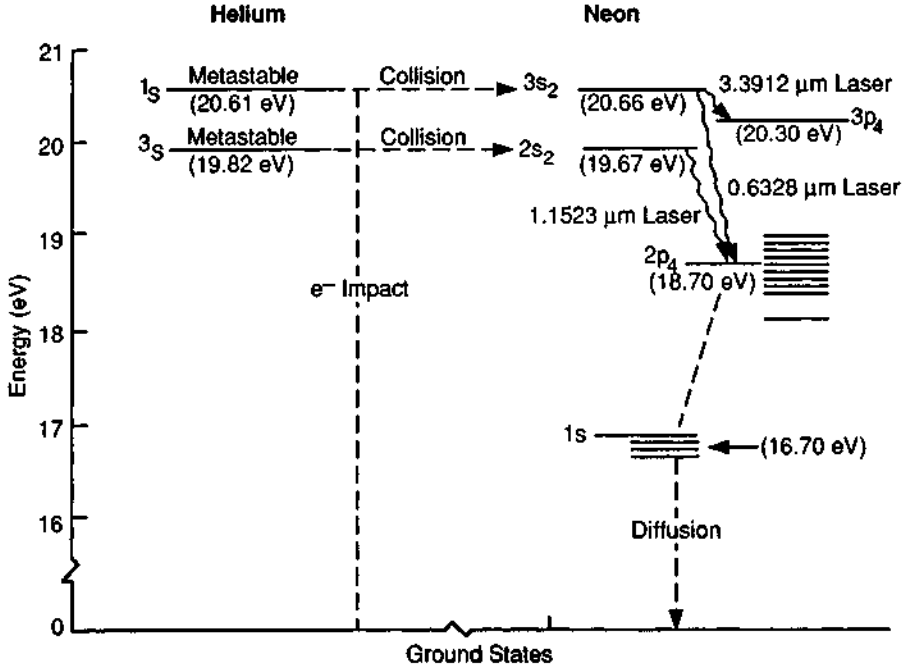


Figure 1-10 Energy levels relevant to the operation of the helium–neon laser, with the spectroscopic designations indicated. The paths for excitation and laser emission are shown.

been burned in the line." The power output is limited by the fraction of atoms available. Therefore, homogeneous lines are desirable. In general, Lorentzian lines are homogeneous and Gaussian lines are inhomogeneous. Specific examples include

Gases: Inhomogeneous

Ruby at room temperature: Homogeneous

Ruby at cryogenic temperature: Inhomogeneous

Nd:YAG: Homogeneous

Semiconductor: Homogeneous

F. Population Inversion

When light with intensity $I(z)$ travels in the z direction through a material with densities of atoms N_1 and N_2 in two energy levels satisfying Equation (1.7), and numbered so that level 2 lies higher than level 1, the light intensity decreases because of absorption by the atoms in level 1. The decrease is given by

$$I(z) = I(0)\exp(-\alpha z) \tag{1.8}$$

where α is the absorption coefficient and is given (for a Gaussian line shape) by

$$\alpha = \left(\frac{c^2}{4\pi f^2 \tau}\right)\left[\frac{\ln 2}{\pi}\right]^{1/2}\left(\frac{N_1}{\Delta f}\right) \tag{1.9}$$

where τ is the fluorescent lifetime of the material and the other notation is as before. Note that c is the velocity of light in the material, not the velocity of light in vacuum. A slightly different expression is applicable for a Lorentzian line, but α is still proportional to $N_1/\Delta f$.

Amplification by stimulated emission increases the light intensity according to

$$I(z) = I(0)\exp(+gz) \tag{1.10}$$

where g, the gain coefficient, is

$$g = \left(\frac{c^2}{4\pi f^2 \tau}\right)\left[\frac{\ln 2}{\pi}\right]^{1/2}\left(\frac{N_2}{\Delta f}\right) \tag{1.11}$$

for a Gaussian line. We note the similarity of the forms for α and g. This means that absorption and stimulated emission are inverse processes. The net effect of passage of the light through the material is

$$I(z) = I(0)\exp\left\{\left(\frac{c^2}{4\pi f^2 \tau}\right)\left[\frac{\ln 2}{\pi}\right]^{1/2}\left(\frac{1}{\Delta f}\right)(N_2 - N_1)z\right\} \tag{1.12}$$

If $N_2 > N_1$, the gain via stimulated emission will exceed loss via absorption, the exponential will be positive, and the light intensity will increase as it travels through the material. If $N_1 > N_2$, the light intensity will decrease. For a laser to operate, one requires $N_2 > N_1$. This situation is called a "population inversion." The population inversion is attained simply by having more electrons in a high-lying energy level at which fluorescent emission starts than in the lower-lying energy level at which it terminates. Without the population inversion, according to Equation (1.12) there will be a net absorption at frequency f. With the population inversion, there will be light amplification by stimulated emission of radiation.

A population inversion is not the usual situation. For a material in thermal equilibrium at temperature T, one has

$$\frac{N_2}{N_1} = \exp\left[\frac{-(E_2 - E_1)}{kT}\right] \tag{1.13}$$

where k is Boltzmann's constant, equal to 1.38×10^{-16} erg/deg. This equation means that at thermal equilibrium, $N_1 > N_2$. To achieve laser operation, one must upset the thermal equilibrium in some way so as to produce the nonequilibrium situation of a population inversion.

We now consider the so-called threshold condition, which defines the minimum population inversion required for laser operation in a particular system. There are inevitable sources of loss associated with the laser. One loss is output of light energy through the two end mirrors of the laser, with reflectivities R_1 and R_2. (We assume here that the laser has two end mirrors; we shall consider this further in the discussion of resonant structures.)

Another source of loss involves imperfections in the system, such as inhomogeneities in a solid laser material. Such losses may be characterized by a loss coefficient β such that the decrease in light intensity as the light travels a distance z is given by a factor of $exp(-\beta z)$. According to Equation (1.12), the change in light intensity after a round-trip through the active medium between two mirrors separated by distance D involves a factor

$$R_1 R_2 \exp(-2\beta D) \exp\left\{2D\left(\frac{c^2}{4\pi f^2 \tau}\right)\left[\frac{\ln 2}{\pi}\right]^{1/2}\frac{(N_2 - N_1)}{\Delta f}\right\} \tag{1.14}$$

For a net increase in light intensity, this factor must be greater than unity. Taking logarithms, one obtains

$$N_2 - N_1 > \left(\frac{4\pi f^2 \tau \Delta f}{Dc^2}\right)\left(\frac{\pi}{\ln 2}\right)^{1/2}(\beta D - \ln R_1 R_2) \tag{1.15}$$

as the threshold condition, that is, the minimum population inversion required to sustain laser action in a Gaussian line for a medium with the given properties. For a Lorentzian line, the expression will differ in some inconsequential details. One sees that a small value of Δf, that is, a narrow spectral linewidth, is desirable to make it easy to produce the required population inversion. Equation (1.15) is important because it defines the threshold at which the laser will operate. One must take steps to supply the minimum population inversion specified by this equation.

The population inversion is produced by some method that supplies energy to electrons and raises them to high-lying energy levels. This process is commonly referred to as "pumping." Common methods of supplying the energy for the population inversion are optical excitation with a light source or excitation via collisions with electrons in a gas. We will discuss this in more detail shortly.

It is impossible to produce a population inversion by optical pumping in an electronic system with only two energy levels. Thus, one uses materials with three or four levels. (The actual number of energy levels in any laser material will be much greater than four, but only three or four will be directly involved in the laser operation.) Laser materials are specified as being "three-level systems" or "four-level systems." The distinction is that in a three-level system, the terminal level for the fluorescence is the ground level (the level with the lowest energy). In a four-level system, the terminal level lies above the ground level. The situation is sketched in Figure 1-11. Consider first the three-level system. One may produce the population inversion by raising electrons to high-lying energy levels (pumping) with an auxiliary light source (the pump). Light with frequency $f_3 = (E_3 - E_1)/h$ is absorbed to excite electrons from level 1 to level 3. Then, a very fast radiationless transition (accomplished by thermal vibrations of the atoms) will drop the electrons to level 2. The difference in energy between levels 3 and 2 appears as heat. Stimulated emission occurs between levels 2 and 1 at frequency $f_2 = (E_2 - E_1)/h$. If one supplies substantial power at frequency f_3, the transition rate from level 1 to level 3 will be large and one can have $N_2 > N_1$, so that the population inversion is achieved. But the three-level system has the drawback that level 1 is the ground level, which normally has

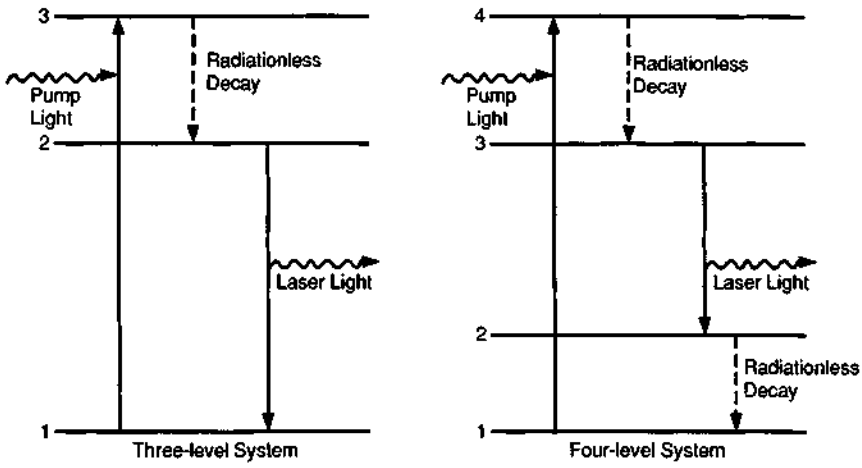


Figure 1-11 Schematic diagram of excitation processes and laser emission for three- and four-level laser systems.

almost all the electrons residing in it. More that half of all the electrons must be moved to level 2, and so intense pumping is needed.

A four-level system removes this constraint. Radiationless transitions from level 4 to level 3 and from level 2 to level 1 may be very fast. Level 2 will remain empty, or almost so. Then, to produce a population inversion, one needs only to get a small number of electrons into level 3, not half of all the electrons. Thus, four level systems are desirable because they require less intense pumping and offer lower threshold excitation energies. Also, there is less heat generation via the radiationless transitions, so that cooling requirements are eased. Most useful lasers are based on four-level systems.

In practice, there are three main methods of supplying the energy needed for the population inversion: optical pumping, excitation by collisions with energetic electrons, and resonant transfer of energy by collisions with other molecules.

1. OPTICAL PUMPING

Optical pumping involves use of an auxiliary source of light that is absorbed by the active medium to raise electrons to high-lying energy levels. This has already been mentioned in reference to Figure1-11.

The ruby laser (see Figure 1-9) employs optical pumping to produce a population inversion. Light in the green portion of the spectrum (520–580 nm) is absorbed by electrons in the ground state, raising them to the highest level shown for this three-level system. Radiationless decay to the upper level for the 694.3 nm fluorescent emission then follows. With intense pumping, the ground state will be depopulated, a population inversion will be produced, and laser action can occur.

One of the most common light sources for optical pumping is a flashlamp similar to those used in photography. The flashlamps used in lasers are glass or quartz tubes filled with gas at low pressure. Commonly, xenon gas is used, although in some

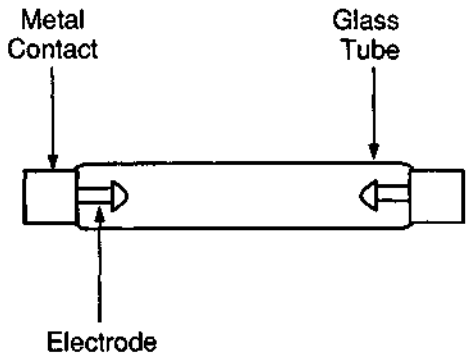


Figure 1-12 Structure of a linear flashlamp.

cases other noble gases, like krypton, are employed to produce light that gives a better match to the absorption spectrum of the active material. The lamps have internal electrodes and can be prepared in a variety of shapes. Figure 1-12 shows a typical linear lamp. Helical and U-shaped lamps have also been used. The range of performance for flashlamps is very broad. Electrical input energy, lamp size, and duration of the light pulse can be varied over a wide range.

When an electric current is discharged through the lamp, a high-temperature plasma is formed. The plasma emits radiant energy over a wide spectral range. The radiation generally has a complicated spectrum, containing both line and continuum components. Within reasonable limits, it is possible to vary the gas in the flashlamp and the operating conditions to match some portion of the emission spectrum to the absorption spectrum of the laser material.

Figure 1-13 shows a typical circuit including an energy storage capacitor C and a high-voltage trigger pulse which is applied externally to the flashlamp. The discharge in the flashlamp is initiated by applying a trigger pulse to the electrodes, causing a spark streamer to form within the flashlamp. Once the flashlamp has been triggered, the capacitor can discharge through the ionized gas. The series inductance L helps shape the current pulse through the lamp. It also reduces ringing or reverse flow of current, which is harmful to the lamp lifetime. The lifetime of flashlamps will be considered in Chapter 6.

Flashlamps are used for optical pumping of pulsed lasers. Arc lamps, which also consist of a gas-filled tubular structure with metal electrodes, operate continuously and are employed for optical pumping of continuous lasers

Another source for optical pumping is the semiconductor diode laser. These lasers are especially useful for pumping solid state lasers, like Nd:YAG. Compared with flashlamps, diode lasers offer a better match of the output spectrum to the absorption of the material. This leads to increased efficiency and results in a smaller, lighter system. Diode laser pumping is becoming very common for solid state lasers, as we shall discuss in Chapter 3.

Optical pumping is often used with solid state lasers, like ruby and Nd:YAG. In these materials, there is no possibility of electrical current flow through the material

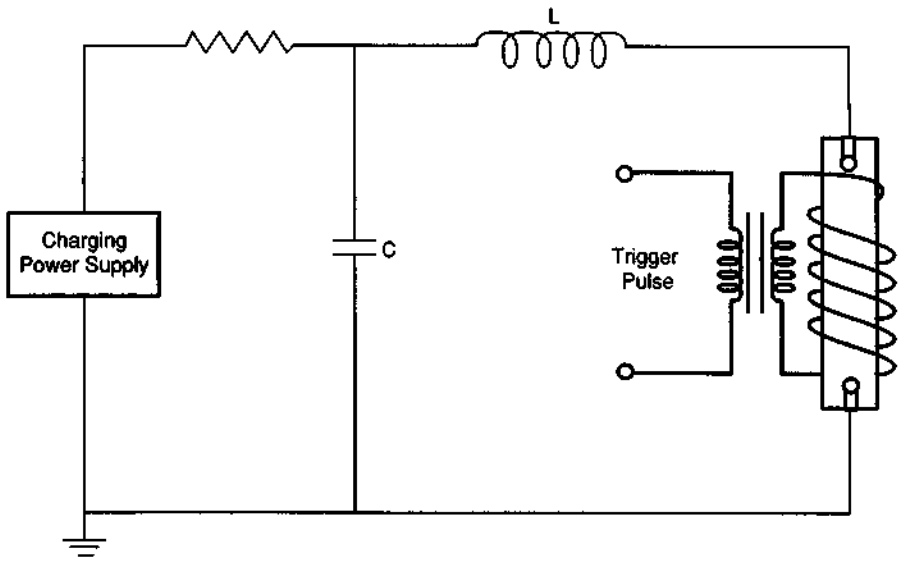


Figure 1-13 A typical circuit for operation of a flashlamp, including energy storage capacitor C and series inductance L.

with resultant excitation through collisions with electrons. But in a gas, electrical discharges are possible, and atoms can be excited via collisions with energetic electrons.

2. *EXCITATION BY ELECTRON COLLISIONS*

Excitation of high-lying energy levels via collision of a molecule with an energetic electron is shown schematically in Figure 1-14. Electrons are injected into the material and collide with atoms in the ground state, raising the electrons to energy level 4. Rapid radiationless decay to level 3 follows. The population inversion is produced between levels 3 and 2. (Level 2 is originally empty.)

A specific example involves an electrical discharge through a gas that constitutes the active medium. The structure may be similar to a neon advertising sign. Electrical current flows between electrodes in the gas. Electrons in the current flow collide with gas molecules and transfer their kinetic energy to excitational energy of the gas molecules.

To produce the required population inversion, the upper level 3 must be populated more than level 2 by the collisional process. This implies a suitable balance between the electron velocity spectrum, the cross sections for transferring energy to the different energy levels, and the rates at which the different levels decay. It may be difficult to find suitable combinations such that the threshold condition (Equation 1.15) will be satisfied. Because of this difficulty, lasers do not usually have a single gas with electronic collisions populating the upper level directly. More often, gas lasers employ a variation in which the electrons populate an upper level in one gas and the energy is transferred in resonant collisions to a second molecule.

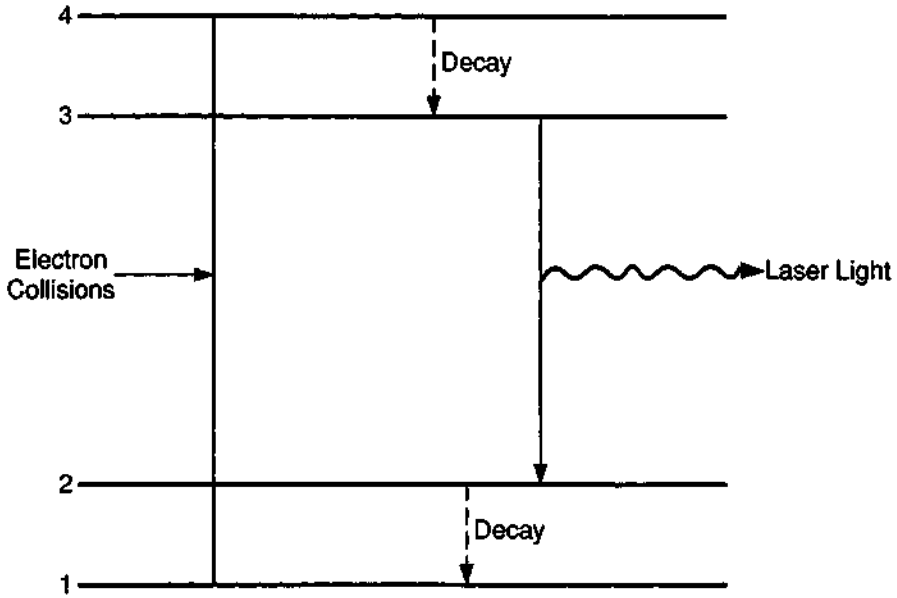


Figure 1-14 Schematic diagram for excitation and laser emission for direct electronic excitation in a four-level laser system.

3. RESONANT TRANSFER OF ENERGY

Gaseous laser materials are commonly excited via a two-step process in which energetic electrons in an electrical discharge collide with molecules and transfer energy to their energy level system, with a subsequent collisional transfer of energy from the first molecule to the gaseous active medium. The method requires two gases that have high-lying energy levels with a near coincidence of energy values. Examples are He and Ne (see Figure 1-10) or CO_2 and N_2. Electrons excite one of the gases to a high-lying energy level in a collision. The excited gas molecules then collide with molecules of the second gas and transfer the excitation energy to it. The molecules of the second gas then are in an upper energy level with the lower energy levels empty, so that the required population inversion is established.

Specifically, in the He–Ne laser, helium atoms are excited by energetic electrons. The excitation energy is transferred to the neon atoms, which constitute the active medium. Laser operation then proceeds using the population inversion established between energy levels in the neon. Although laser operation in pure neon can occur via direct electronic excitation of the neon, addition of helium increases the population inversion and greatly enhances the laser action.

Similarly in the CO_2–N_2 laser (usually simply called the CO_2 laser), the N_2 is excited in electronic collisions and then passes its energy to a coincident level in the CO_2, which produces the laser operation.

The process requires that the two energy levels in the two molecules have almost identical energies. Hence, the transfer of energy is called resonant.

We also note that energy is required to produce the population inversion. The electrical discharge consumes energy, so that the laser output is always less that the energy supplied. Thus, most lasers operate at relatively low efficiency (defined as the ratio of the radiant energy output to the electrical energy input).

We have considered some of the common methods of establishing a population inversion in an active material. Other methods, such as gasdynamic processes, chemical processes, and inversion at a semiconductor junction, will be described later with reference to specific lasers.

G. Resonant Cavity

The final requirement for a laser is some sort of mirror structure to form a resonant cavity. The resonant cavity, or optical resonator as it is sometimes called, is formed by placing mirrors at the ends of the active medium. The mirrors are perpendicular to the axis along which the laser light travels. The earliest lasers used flat mirrors, but it was soon found that improved stability could be achieved with spherical mirrors. Slight misalignment or vibration in a laser with flat mirrors could cause the light to "walk off" the axis. Today, most lasers use spherical mirrors with long radii of curvature because they are less sensitive to alignment.

The resonant cavity generally is much longer than it is wide. This configuration maintains particular configurations of the electromagnetic field with low loss, as we shall describe futher in Chapter 2. The loss is compensated for by the gain from stimulated emission as the light passes through the active medium.

Usually one of the mirrors will have high reflectivity, as close to 100 percent as possible. The other mirror has some transmission, so that light may emerge from the laser. The reflectivity of this mirror is less than 100 percent, with the deviation from 100 percent depending on the amount of gain that can be supplied by the active medium. The departure from 100 percent represents a source of loss.

In our discussion of resonant cavities, we first assume that the resonant cavity consists of two parallel mirrors, and later we consider other arrangements.

To sustain laser operation, one usually must provide a resonant cavity so that the light will make many passes through the active medium. Such an arrangement is needed so that the light can effectively travel a long path in the medium and allow the light intensity to increase via stimulated emission There are a few materials that have very high gain coefficient and do not require a long path in the material. These materials can dispense with the resonant cavity; such lasers are called superradiant. But the vast majority of lasers do use a resonant cavity.

Figure 1-15 shows schematically a resonant cavity formed by two mirrors on the end of a solid laser rod excited by a flashlamp. Here, the mirrors are reflecting coatings on the flat and parallel ends of the rod. Of course, many other arrangements are possible, such as separate mirrors removed from the ends of the rod.

Figure 1-15 shows all the elements needed for the laser: the active material, the means for producing the population inversion (the flashlamp plus a reflector to collect the light and concentrate in on the rod), and the resonant cavity (the mirrors). For the electromagnetic wave to be resonant in the cavity defined by the mirrors, the

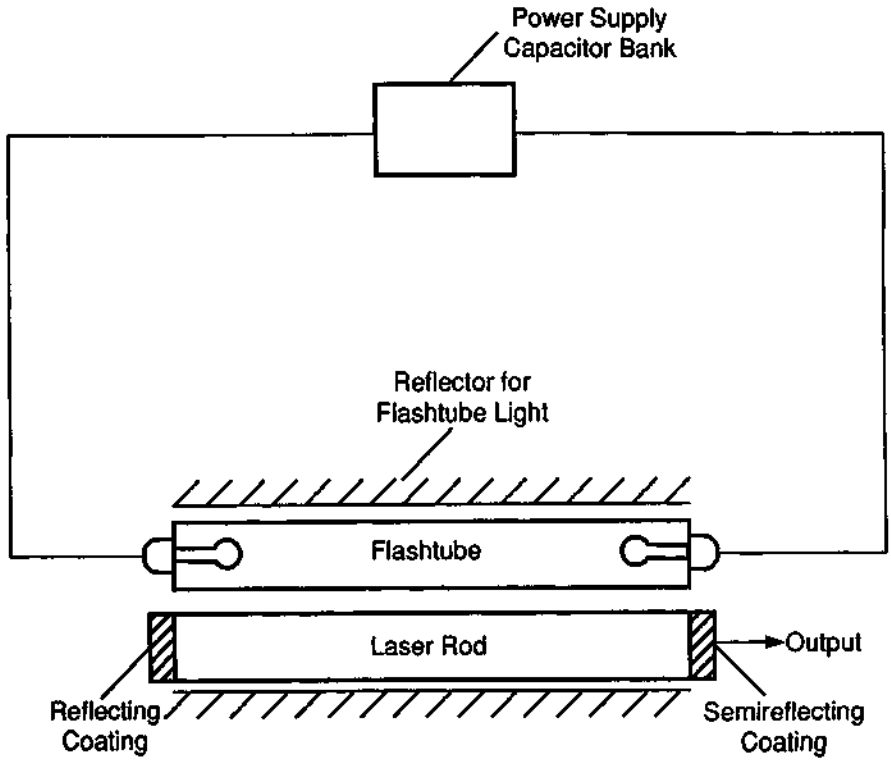


Figure 1-15 An optically pumped solid laser.

field distribution in the cavity must replicate itself after round-trip reflection by the two mirrors. This condition is satisfied only for certain discrete frequencies, which we denote f_{mnq}, where m, n, and q are integers. These resonant frequencies are given by

$$f_{mnq} = [q + (m + n + 1)\cos^{-1}(g_1 g_2)^{1/2}]\frac{c}{2\pi nD} \tag{1.16}$$

where

$$g_1 = 1 - \frac{D}{R_1}, \qquad g_2 = 1 - \frac{D}{R_2} \tag{1.17}$$

with n the index of refraction, D the distance between the mirrors, and R_1 and R_2 the radii of curvature of the mirrors. This equation is simplified for the common case of the mirrors being identical and having long radius of curvature R. In this case, $g_1 = g_2$ and $D/R \ll 1$. Then Equation (1.16) simplifies to

$$f_{mnq} = \left[q + (m + n + 1)\left(\frac{2D}{\pi R}\right)\right]\left(\frac{c}{2nD}\right) \tag{1.18}$$

In these equations m and n are small integers, of the order of magnitude of unity. They define the transverse mode structure, which we will discuss in Chapter 2. On the other hand, q is a large integer, perhaps lying in the range 10^5 to 10^6. Then we may simplify further. As an approximation, the resonant frequencies depend mainly on q and are given by

$$f = \frac{qc}{2nD} \tag{1.19}$$

In terms of wavelength, one has

$$q\left(\frac{\lambda}{2}\right) = nD \tag{1.20}$$

This expression appears plausible: if one wishes to have a condition of resonance so that the light bouncing back and forth between the mirrors is in proper phase for buildup in the stimulated emission process, the mirrors should be separated by an integral number of half-wavelengths of the light. Equation (1.20) is the equation commonly quoted for the resonant wavelengths.

Usually, there are many combinations of λ and q that satisfy this equation and for which λ lies within the fluorescent linewidth of the active medium. These are called cavity modes. The output of the laser does not consist of a continuum of wavelengths covering the entire fluorescent line of the active medium. Rather, laser operation occurs only at these discrete wavelengths satisfying Equation (1.20) for some integer q. Figure 1-16 shows the situation. The fluorescent line shape is shown, with the positions of the resonant cavity modes indicated. They are evenly spaced in wavelength, with separation $\lambda^2/2D$.

These modes represent specific wavelengths at which a given laser (specified by the active medium and the mirror separation) can operate. They are commonly called longitudinal modes.

These modes are similar to the resonant modes of a microwave cavity, except the wavelength is of the same order of magnitude as the cavity length for the microwave case. For microwaves, the integer q would be of the order of unity. For visible and infrared light, the geometrical structure is much larger than the wavelength and q can easily be of the order of 10^5 or 10^6.

The presence of the longitudinal modes can make the spectral content of the laser output complicated. Usually, the laser output will contain a number of different frequency components, as Figure 1-16 indicates. Thus, the frequency spectrum consists of several discrete, separated frequencies, each arising from the presence of a different longitudinal mode satisfying Equation (1.20). We note that the wavelength difference between longitudinal modes is very small compared with the operating wavelength of the laser. Thus, the laser output is almost monochromatic despite the presence of the different wavelength components.

A complete discussion of resonant modes of the cavity should also include the spatial or transverse mode structure. This will be deferred until Chapter 2, in the description of the spatial profile of the laser beam.

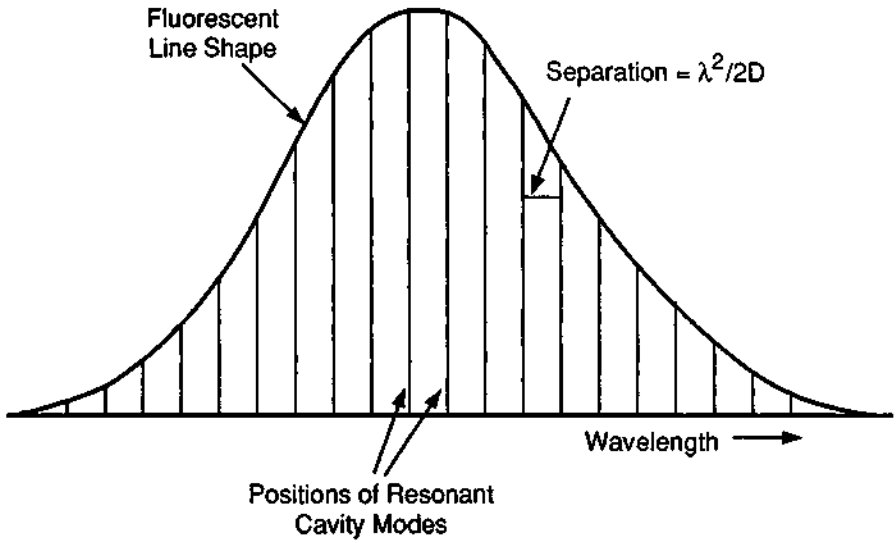


Figure 1-16 Positions of the resonant modes in a laser cavity with length D and wavelength λ.

Now let us consider the growth of the radiation field between the mirrors. Light is reflected back and forth by the mirrors, and it is amplified by stimulated emission on each pass through the active medium. The increase of the light is shown schematically in Figure1-17. Parts (a) and (b) show the original unexcited material, followed by excitation via optical pumping. Part (c) shows fluorescence, followed by the start of stimulated emission. Parts (d) and (e) show subsequent growth of the light intensity by stimulated emission. The ends of the laser are flat and parallel and allow the light to pass back and forth many times through the active medium. Some fluorescent light will be present because of fluorescent emission from excited electrons. Some of this fluorescent light will travel in a direction other than the axis of the laser. This light will reach the side of the laser and will be lost. A fraction of the fluorescent light will be aligned with the axis of the resonant cavity. This is the light that will be amplified by the stimulated emission process to form the laser output. At the semitransparent mirror at one end of the laser, some of the light emerges as the useful output of the laser. Some is reflected back for another pass through the active medium and a further buildup in intensity. This light will pass back and forth through the active material many times.

The result of this process is a regenerative cascade in which energy stored in upper energy levels of electrons is rapidly converted into an intense, highly collimated beam of light that emerges through one of the mirrors. Although this discussion might seem to indicate that the growth of light intensity and extraction of light from the mirrors is a succession of separate bounces between the mirrors, it is, of course, a continuous process.

So far, we have assumed that the resonant cavity consists of two plane parallel mirrors. Because plane parallel mirrors require exact angular alignment if the light is to make many bounces back and forth between them, the use of spherical mirrors,

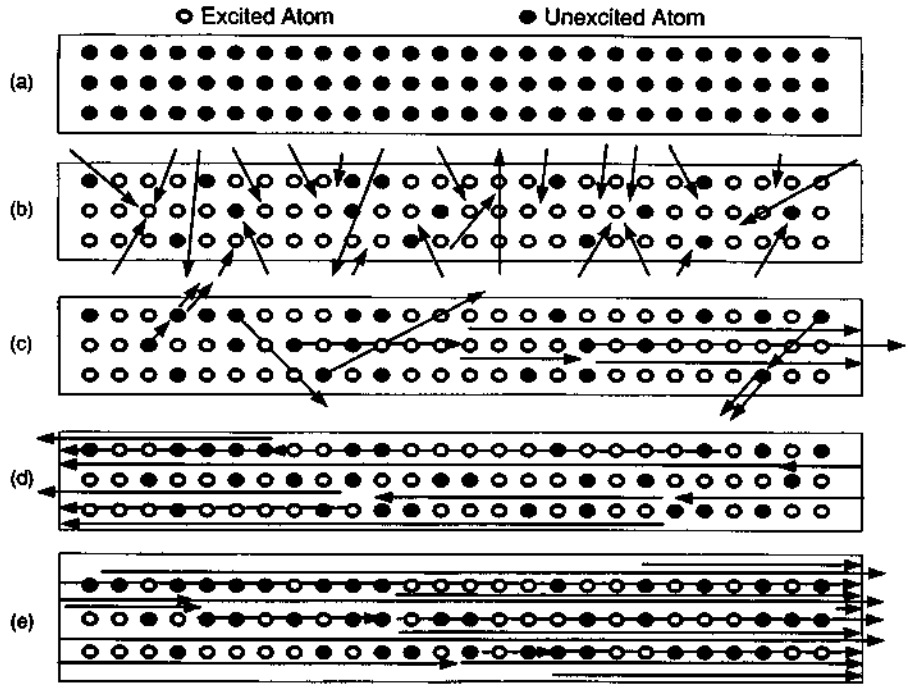


Figure 1-17 Excitation and stimulated emission in a solid laser rod. (a) Initial unexcited rod. (b) Excitation by pump light. (c) Fluorescence and start of stimulated emission. (d) Growth of stimulated emission preferentially along the rod axis. (e) Continued amplification by stimulated emission and output from the end of the rod.

which are less sensitive to angular alignment, has become common. Resonant cavities formed by spherical mirrors can have reentrant paths for the light such that the light will make many passes through the active medium. If R_1 and R_2 represent the radii of curvature for the two spherical mirrors, and D is the distance between them, one can have stable laser operation for many combinations of R_1, R_2, and D.

Mirror configurations may be judged on two criteria: stability and filling of the active medium by the light. Stability means that light rays bouncing back and forth between the mirrors will be reentrant; that is, the electric field distribution associated with the light rays will replicate itself after a round-trip between the mirrors. Good filling of the active medium means that the spatial profile defined by the light rays fills all the volume of the active medium contained between the mirrors.

Plane parallel mirrors have good filling, because the light rays may fill the entire volume between the mirrors. If they are perfectly aligned, the light rays are reentrant, but if the mirrors are even slightly misaligned, the light rays will walk off the edges of the mirrors after a number of reflections. Thus, plane parallel mirrors have marginal stability. Some combinations of spherical mirrors have good stability, so that if the configuration is not exactly perfect, the light rays will still be reentrant. However, not all spherical mirrors fill the volume effectively. An example is the con-

focal case, for which $R_1 = R_2 = D$. The focal point for each mirror is at the midpoint of the cavity, so that the light rays pass through a narrow waist halfway between the mirrors. They will not make use of the active material that is off the center axis halfway between the mirrors. Thus, a confocal configuration has poor filling of the active material.

The condition for stability is given by

$$0 < \left(\frac{D}{R_1} - 1\right)\left(\frac{D}{R_2} - 1\right) < 1 \tag{1.21}$$

This is illustrated in Figure 1-18, which shows a stability diagram for resonator geometries with the quantities D/R_1 and D/R_2 as the coordinates. Each resonator system is represented as a point in this diagram. Points in the unshaded regions indicate stable resonators, and points in the shaded regions indicate unstable systems.

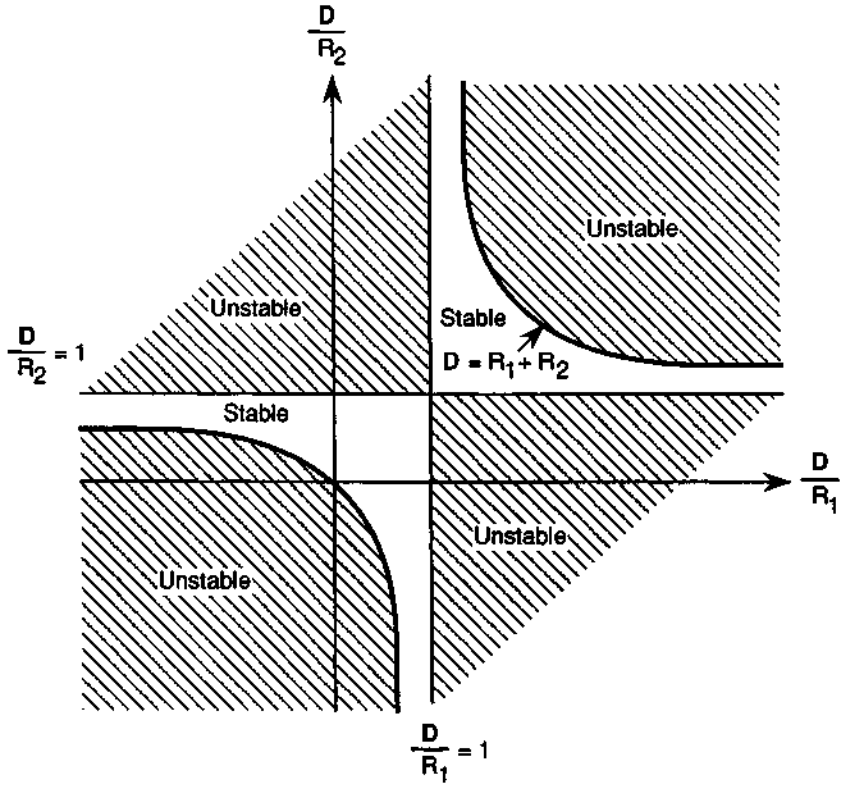


Figure 1-18 Stability diagram for laser resonators consisting of two mirrors with radii of curvature R_1 and R_2 separated by distance D. Each point represents a particular geometry. The unshaded regions represent stable systems; the shaded areas represent unstable systems. (From G. D. Boyd and H. Kogelnik, *Bell System Technical Journal* **41**, 1347 (1962). Copyright © 1962 AT&T. All rights reserved. Reprinted with permission.)

For example, the plane parallel mirror system is represented by the origin, which is on a boundary between stable and unstable regions. The confocal configuration, at the intersection of the lines $D/R_1 = 1$ and $D/R_2 = 1$, is also of marginal stability. Long-radius mirrors ($R_1 \gg D$, $R_2 \gg D$) lie completely within a stable region slightly to the upper right of the origin.

A number of possible mirror configurations are sketched in Figure 1-19. In practice, long-radius mirrors are often used because they fall within a region of good stability in Figure 1-18 and the beam spatial profile fills the active medium reasonably well. Plane parallel mirrors offer good use of the medium, but they have delicate stability. Hemispherical mirrors ($R_1 = D$, $R_2 = \infty$) offer good stability, but the radiation does not fill the medium well. The spherical system ($R_1 = R_2 = D/2$) and the confocal system both have poor stability and do not fill the active medium. An example of an unstable configuration (one of many) is also shown in the figure.

The long-radius configuration represents a good compromise between stability and filling of the active medium. It is the geometry most often used in modern commercial lasers.

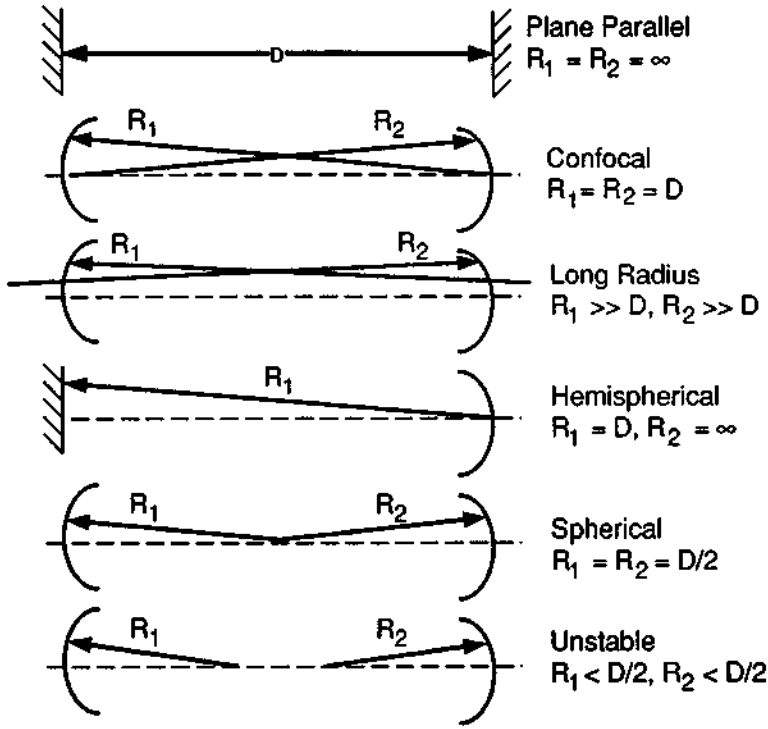


Figure 1-19 A number of possible mirror configurations. R_1 and R_2 are the mirror radii, and D is the mirror separation.

In some conditions, unstable resonators are in fact used in lasers. Such resonators are most useful for lasers with Fresnel number much greater than unity. The Fresnel number N_F is defined by

$$N_F = \frac{a^2}{D\lambda} \tag{1.22}$$

where a is the radius of the laser medium and the other notation is as before. For long, slender lasers, the Fresnel number may be of the order of magnitude of unity, and the stable, long-radius configuration is useful. However, if the radius is large, the Fresnel number may become large compared with unity. The spatial structure of the beam may become complicated, and the beam divergence will be large. (See Chapter 2 for further discussion.) In this case, an unstable resonator may produce a more desirable spatial pattern for the beam.

The laser beam emerges around rather than through the output mirror. A typical unstable resonator configuration is shown in Figure 1-20. In this situation, both mirrors have high reflectivity. The loss of the system tends to be high, so that unstable resonators are useful only with materials that can supply high gain, typically more than 50 percent per pass. Unstable resonators find most use with very high power gas lasers. Figure 1-20 also shows a typical method of bringing the beam out of the laser with an unstable resonator. Unstable resonators have probably been used most often with CO_2 lasers. They are well suited for use with high-power CO_2 lasers, which can have large diameter and high gain per pass. They have been used to produce well-collimated beams of higher quality than would be available with stable resonators.

The reflecting surfaces commonly used to form the mirrors for lasers are evaporated coatings consisting of multiple layers of dielectric materials. The technology of these mirrors will be described in Chapter 5.

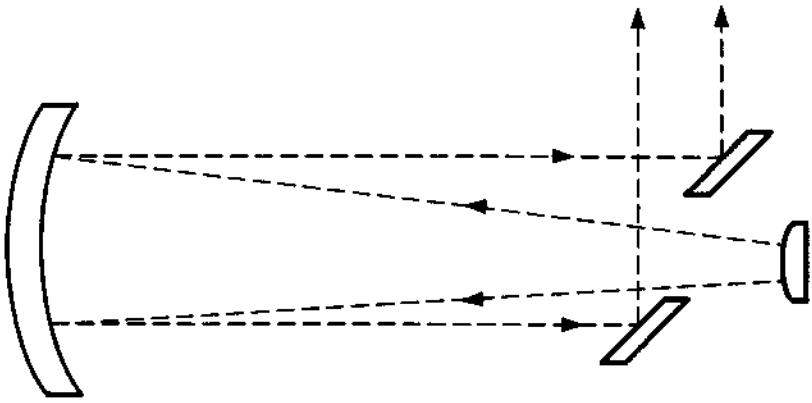


Figure 1-20 An unstable resonator.

Selected References

A. Electromagnetic Radiation

C. L. Andrews, *Optics of the Electromagnetic Spectrum*, Prentice Hall, Englewood Cliffs, NJ, 1960, Chapter 1.
M. Garbuny, *Optical Physics*, Academic Press, New York, 1965, Chapter 1.

B. Elementary Optical Principles

M. Born and E. Wolf, *Principles of Optics*, 6th ed., Pergamon, Oxford, 1980.
F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th ed., McGraw-Hill, New York, 1976.
M. V. Klein, *Optics*, Wiley, New York, 1970.

C. Energy Levels

W. S. C. Chang, *Principles of Quantum Electronics*, Addison-Wesley, Reading, MA, 1969, Chapter 2.
A. E. Siegman, *Lasers*, University Science Books, Mill Valley, CA, 1986.
M. J. Weber, ed., *Handbook of Laser Science and Technology; Vol. I: Lasers and Masers; Vol. II: Gas Lasers*, CRC Press, Boca Raton, FL, 1982.

D. Interaction of Radiation and Matter

M. Garbuny, *Optical Physics*, Academic Press, New York, 1965, Chapter 3.
B. A. Lengyel, *Introduction to Laser Physics*, Wiley, New York, 1966, Chapter 1.
A. Maitland and M. H. Dunn, *Laser Physics*, North Holland, Amsterdam, 1969, Chapters 1 and 3.
J. R. Singer, *Masers*, Wiley, New York, 1959, Chapter 2.
O. Svelto, *Principles of Lasers*, 3rd ed., Plenum, New York, 1989.
A. Yariv, *Quantum Electronics*, 3rd ed., Wiley, New York, 1988.

E. Laser Materials

A. L. Bloom, *Gas Lasers*, Wiley, New York, 1968, Chapter 2.
W. S. C. Chang, *Principles of Quantum Electronics*, Addison-Wesley, Reading, MA, 1969, Chapter 7.
M. J. Weber, ed., *Handbook of Laser Science and Technology, Vol. I: Lasers and Masers: Vol. II: Gas Lasers*, CRC Press, Boca Raton, FL, 1982.

F. Population Inversion

A. L. Schawlow, Optical Masers, *Sci. Am.*, p. 2 (June 1961).
A. L. Schawlow, Advances in Optical Masers, *Sci. Am.*, p. 34 (July 1963).
A. E. Siegman, *Lasers*, University Science Books, Mill Valley, CA, 1986.
O. Svelto, *Principles of Lasers*, 3rd ed., Plenum, New York, 1989.
C. S. Willett, *Introduction to Gas Lasers: Population and Inversion Mechanisms*, Pergamon, Oxford, 1974.
A. Yariv, *Quantum Electronics*, 3rd ed., Wiley, New York, 1988.

G. Resonant Cavity

G. D. Boyd and J. P. Gordon, Confocal Multimode Resonator for Millimeter through Optical Wavelength Masers, *Bell Syst. Tech. J.* **40**, 489 (1961).
A. G. Fox and T. Li, Resonant Modes in an Optical Maser, *Proc. IRE* **48**, 1904 (1960).
A. G. Fox and T. Li, Resonant Modes in a Maser Interferometer, *Bell Syst. Tech. J.* **40**, 453 (1961).
H. Kogelnik and T. Li, Laser Beams and Resonators, *Appl. Opt.* **5**, 1550 (1966).
A. Maitland and M. H. Dunn, *Laser Physics*, North Holland, Amsterdam, 1969, Chapters 1, 4, 5, and 6.

Chapter 2 | Properties of Laser Light

The applications of lasers depend on the unusual properties of laser light, ones that are different from the properties of light from conventional sources. The properties that will be discussed in this chapter include monochromaticity (narrow spectral linewidth), directionality (good collimation of the beam), the spatial and temporal characteristics of laser beams, coherence properties, radiance (or brightness), the capability to focus to a small spot, and the high power levels available. These properties are not all independent: for example, the focusability depends on the collimation to a large extent. Still, it is convenient to discuss these different aspects separately. These characteristics of laser light, which can be very different from light produced by ordinary sources, enable lasers to be used for the practical applications, which will be described later.

A. Linewidth

Laser light is highly monochromatic, that is, it has a very narrow spectral width. The spectral width is greater than zero, but typically it is much less than that of conventional light sources. The narrow spectral linewidth is one of the most important features of lasers. Early calculations [1] indicated that the linewidth could be a small fraction of 1 Hz. Of course, most practical lasers have much greater linewidth.

In discussing linewidth, we must distinguish between the width of one mode of the resonant cavity and the width of the total spectrum of the laser output. As Chapter 1 described, many lasers operate simultaneously in more than one longitudinal mode, so that the total linewidth may approach the fluorescent linewidth of the laser material. Figure 2-1 illustrates the situation. The evenly spaced cavity modes are illustrated in part (a). The frequency spacing is $c/2nD$, with D the distance between the two mirrors and n the index of refraction of the material. For a gas, the index of refraction is close to unity, and the spacing becomes $c/2D$. This spacing is equivalent to a wavelength spacing of $\lambda^2/2D$, where λ is the wavelength. Figure 2-1b shows the gain and the loss as functions of frequency. The gain curve has the same shape as the fluorescent line, and the loss is essentially independent of frequency, at least

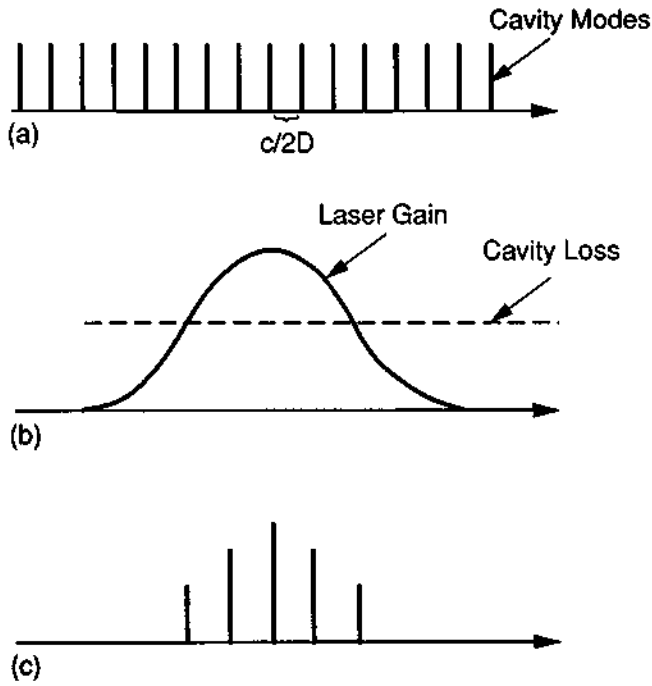


Figure 2-1 Frequency spectrum of laser output. Part (a) shows the resonant modes of a cavity for a gas laser, with c the velocity of light and D the mirror separation. Part (b) shows the gain curve for the fluorescent emission, with the cavity loss, which is relatively constant with respect to frequency, also indicated. Part (c) shows the resulting frequency spectrum, in which all the modes for which gain is greater than loss are present.

over a reasonably small range. All modes for which the gain is greater than loss can be present in the laser output. Thus, the spectral form of the output can be as shown in Figure 2-1c.

Some typical values are shown in Table 2-1 for representative lasers. The fluorescent linewidth varies considerably for different laser materials. The table presents the operating wavelength and frequency of the laser, the fluorescent linewidth of the laser material, and the typical number of longitudinal modes present in the output of the laser. The number of modes is calculated as the fluorescent linewidth divided by the quantity $c/2D$, the spacing between longitudinal modes. In most cases this will be an overestimate, because near the edges of the fluorescent line, the gain will not be high enough to sustain laser operation. Thus, for typical commercial helium–neon lasers, perhaps three or four modes will be present, rather than ten as the calculation indicates. CO_2 lasers present an interesting case, because at low gas pressure the fluorescent linewidth is narrow, of the order of 60 MHz. As the pressure increases to near atmospheric, the fluorescent linewidth broadens. Atmospheric pressure CO_2 lasers may operate in several modes. Neodymium:glass has an unusually wide fluorescent linewidth, and an Nd:glass laser can have hundreds or even thousands of longitudinal modes present in its output. In particular, in mode-locked

Table 2-1 **Linewidth for Common Lasers**

Laser	Wavelength (μm)	Frequency (Hz)	Fluorescent linewidth (MHz)	Typical number of modes present
He–Ne	0.6328	4.7×10^{14}	1700	~10
Argon	0.4880, 0.5145	6.1 & 5.8×10^{14}	3500	~20
CO_2 (low pressure)	10.6	2.8×10^{13}	60	~1
CO_2 (atmospheric pressure)	10.6	2.8×10^{13}	3000	~20
Ruby (300 K)	0.6943	4.3×10^{14}	3×10^5	~200
Ruby (77 K)	0.6934	4.3×10^{14}	1×10^4	~6
Nd:glass	1.06	2.8×10^{14}	6×10^6	~4000

operation (to be described later) the emission spectrum can fill the fluorescent linewidth.

In summary, the total spectral width covered by many practical lasers will be almost as wide as the fluorescent linewidth, often around 10^9 Hz. We note that this is still much smaller than the operating frequency, around 10^{15} Hz. Even a laser with many longitudinal modes present in its output is still almost monochromatic.

Individual longitudinal modes have much narrower spectral width. To provide better monochromaticity, lasers are sometimes constructed so as to operate in only one longitudinal mode.

One method for providing single-mode operation involves construction of short laser cavities, so that the spacing between modes ($c/2D$) becomes large and gain can be sustained on only one longitudinal mode. Other longitudinal modes will fall outside the fluorescent line. This is illustrated in Figure 2-2. This approach is most commonly applied in helium–neon lasers. Commercial single-mode helium–neon lasers have spectral widths of the order of 10^7 Hz, much reduced from the 10^9 Hz range of multimode lasers. But this reduction is achieved at a sacrifice in output power. The short cavity limits the power that can be extracted.

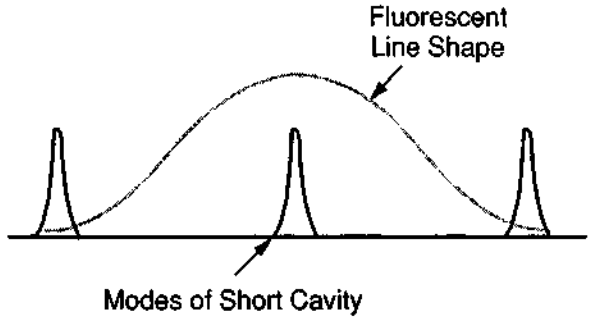


Figure 2-2 Relation of line shape and resonant modes of a short cavity. Only one mode lies within the width of the line, which allows laser operation in a single longitudinal mode.

A second method for obtaining single-mode operation involves use of multiple mirrors, as illustrated in Figure 2-3a. This defines cavities of two different lengths, D_1 and D_2. The length of the short cavity D_1 is chosen so that only one mode lies within the fluorescent linewidth. The laser must operate in a mode that is simultaneously resonant in both cavities. This is illustrated in Figure 2-3b. The lengths of the two cavities must be adjusted carefully, so that there is an overlap of two resonant modes. This places stringent requirements on mirror position and stability. However, this arrangement allows operation of single-longitudinal-mode lasers with higher power. It is used most often with solid state lasers.

Another method for providing a long cavity and a short cavity involves insertion of a device called an etalon into the laser cavity. The etalon is a short resonant cavity, which may be formed from a piece of glass with its two faces polished to a high degree of parallelism. Figure 2-4 shows how an etalon may be inserted into a laser cavity. The etalon generally is tilted with respect to the axis of the cavity in order to provide tuning, as we will discuss. The surfaces of the etalon form a short cavity.

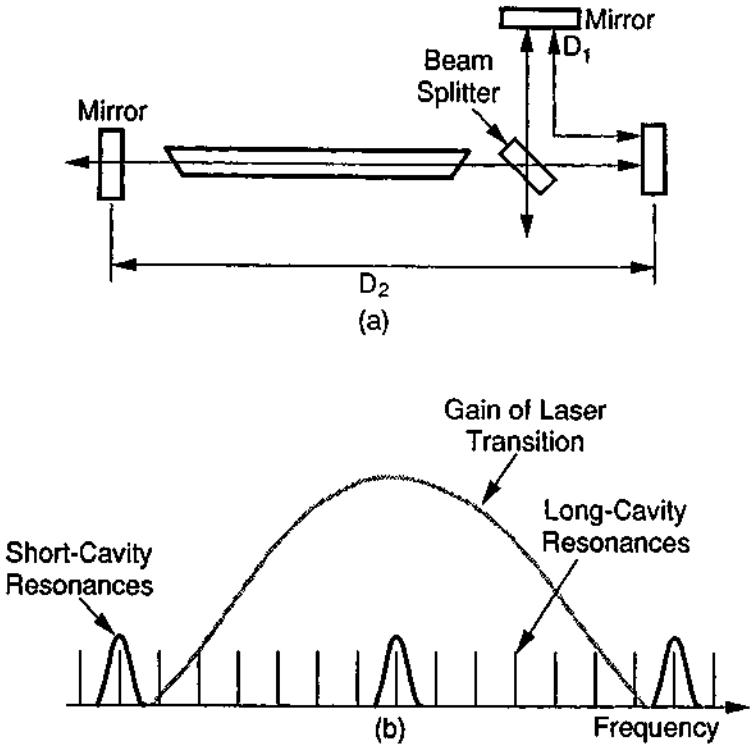


Figure 2-3 Three-mirror method for obtaining operation in a single longitudinal mode. Part (a) shows the positioning of the mirrors so as to define cavities of two different lengths, D_1 and D_2. Part (b) shows the gain of the laser transition superimposed on the longitudinal mode spectra of the long cavity and the short cavity. The laser can operate in only the one mode that is under the gain curve and that is simultaneously a mode of both cavities.

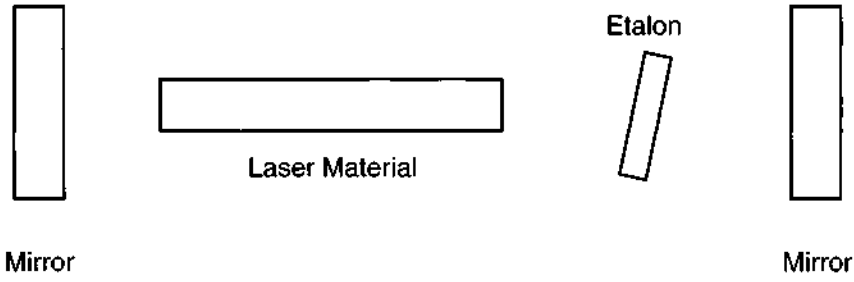


Figure 2-4 Laser cavity with an etalon.

The etalon produces loss at wavelengths that are not modes of its cavity. The longitudinal modes in the laser output must be simultaneously modes of the etalon and of the cavity that defines the laser. Generally, only one mode of the laser will satisfy this condition. This forces the laser to operate in only one of its longitudinal modes. The use of an intracavity etalon is perhaps the most common method for producing single-mode operation.

Usually the etalon is mounted so that its surfaces are tilted with respect to the axis of the laser. As the etalon is rotated, its effective length along the axis changes. The allowed wavelength then changes slightly, so that it is possible to tune the wavelength of the laser by a small amount within the gain curve of the laser material.

Another method for forcing the output of a pulsed laser into a single longitudinal mode is injection seeding. One directs the output of a small, stable, single-frequency laser, called the seed laser, into the laser cavity of the larger laser, which provides the output pulse. The seed laser is usually a continuous laser. The frequency of the seed laser must lie within the linewidth of the larger laser. When the pulse of the larger laser develops, it occurs at the longitudinal mode that is closest to the frequency of the seed laser. This longitudinal mode builds up rapidly by stimulated emission and saturates the gain of the laser before the other longitudinal modes have a chance to develop from noise. Thus, the output of the pulsed laser will be at a single frequency, rather than in a number of longitudinal modes at different frequencies.

Even when operating in a single longitudinal mode, a laser still has a finite spectral linewidth. Because the width of a single mode is much less than the intermode spacing, it will be much smaller than for multimode operation.

Further, the operating frequency of a laser is subject to some random drift. Thermal fluctuations that cause the cavity length to vary are a major contributor to frequency drift. Frequency stabilization may be obtained by vibrational isolation, by temperature stabilization (for example, by operating the laser in an oven with controlled temperature), and by servo control of the mirror spacing according to output power.

In many practical cases, the width of a longitudinal mode is dominated by variations in the length of the laser cavity caused by vibrations and temperature variations. The first steps for reducing this contribution to linewidth are the use of

materials with small thermal expansion coefficients, careful temperature control, isolation from mechanical and acoustic disturbances, and stabilization of gas-laser power supplies to maintain a constant index of refraction in the gas. With such care, single-mode lasers with a short-term frequency stability around 1 kHz are possible.

Improved long-term stability is attained by servo control of the output with respect to some stable reference. For example, the output power of a single-longitudinal-mode laser has a local minimum when the length of the cavity is adjusted so that the laser frequency lies at the center of the spectral line. This local minimum is called the Lamb dip. The mirror spacing is varied by piezoelectric transducers, which are servo controlled to center on the Lamb dip. More elaborate control is possible through use of saturated absorption in a molecular resonance line. This method has apparently been used only for experimental devices, but it has reduced the fractional frequency to a few parts in 10^{13} over 1 sec periods.

One stable gas laser was constructed in a wine cellar in order to achieve very good isolation from vibration [2]. This work achieved a short-term frequency around 20 Hz. The frequency drift was as small as a few tens of hertz per second over several minutes. The device was an experimental model, which has not been duplicated commercially, but it does represent perhaps the best results for frequency stability in a source of electromagnetic radiation.

Commercial single-mode helium–neon lasers are available that have the temperature stabilization and Lamb dip servo control just described. Direct measurements of the linewidth of such lasers are not available. In these commercial lasers, there is a lower limit for the frequency stability around 2 kHz set by the servo dithering on the Lamb dip. Thermal and vibrational instabilities may increase this to around 1 MHz for short periods and to perhaps ±3 MHz over a day. Thus, a reasonable estimate for the frequency stability easily available in a laser is a few parts in 10^9 of the operating frequency.

B. Collimation

One of the most important characteristics of laser radiation is the highly collimated, directional nature of the beam. Collimation allows the energy carried by a laser beam to be easily collected and focused to a small area. For conventional light sources, from which the light spreads into a solid angle of 4π steradians (sr), efficient collection is almost impossible. The small beam divergence angle of a laser means that collection is easy, even at fairly large distances from the laser.

The beam divergence angle θ of a laser is small, but it is not zero. The minimum possible value for the beam divergence arises from the fundamental physical phenomenon of diffraction. This ultimate lower limit is not an engineering limit, which could be reduced by improved optical design. Rather, it arises from a basic physical limitation.

Consider the propagation of a beam with a Gaussian profile (see the next section for the definition of a Gaussian beam). The parameters of most interest in describing the propagation characteristics are the beam radius $w(z)$ at any position z along

the axis of the beam and the radius of the phase front $R(z)$. Figure 2-5 illustrates the propagation of a Gaussian beam, starting from a beam waist of radius w_0. The beam waist is the position of minimum beam diameter, at the position where the phase front becomes plane. The beam radius $w(z)$ and phase front radius $R(z)$, as functions of position z, are given by

$$w(z)^2 = w_0^2\left[1 + \left(\frac{\lambda z}{\pi w_0^2}\right)^2\right] \tag{2.1}$$

$$R(z) = z\left[1 + \left(\frac{\pi w_0^2}{\lambda z}\right)^2\right] \tag{2.2}$$

For most practical lasers, the beam waist lies within the laser cavity. The beam waist w_0 is given by

$$w_0 = \left(\frac{\lambda b}{2\pi}\right)^{1/2} \tag{2.3}$$

where

$$b = \frac{2[D\,(R_2 - D)(R_1 - D)\,(R_1 + R_2 - D)]^{1/2}}{(R_1 + R_2 - 2D)} \tag{2.4}$$

with mirrors having radii of curvature R_1 and R_2 separated by distance D. For many practical cases, the beam waist will be of the same order of magnitude as the radius of the aperture of the laser. The radius of the laser cavity is chosen so that the beam waist will nearly fill the active medium. For example, with gas lasers, the cavity is designed so that the beam almost fills the tube containing the gas.

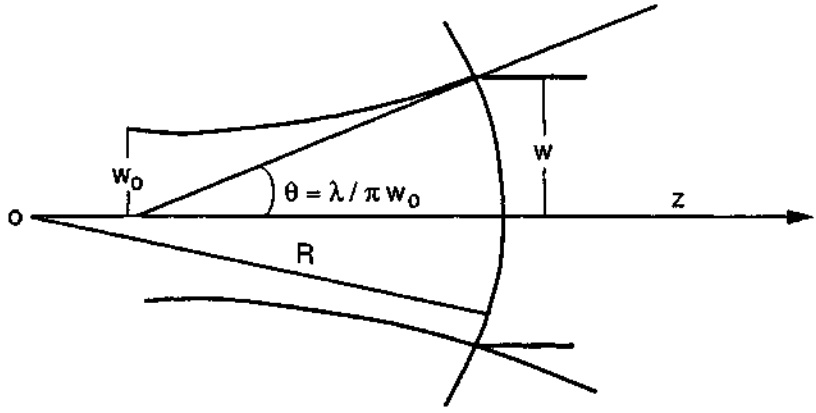


Figure 2-5 Gaussian beam propagation starting from a waist w_0.

For a Gaussian laser beam, the intensity distribution remains Gaussian in every beam cross section as the beam propagates in the z direction. From Equation (2.1), one can see that for large values of z, far from the laser, the beam expands linearly, with a far-field divergence angle θ given by

$$\theta = \frac{\lambda}{\pi w_0} \tag{2.5}$$

if one assumes that θ is reasonably small. In the case where the beam waist fills the aperture of the laser, of diameter D, the approximate relation will be

$$\theta = \frac{2\lambda}{\pi D} \tag{2.6}$$

Note that this is the half-angle of the beam divergence, and that it is in radian measure.

As an example, consider a helium–neon laser with wavelength 0.63 μm and a Gaussian beam profile that fills the inside of a tube with a diameter of 0.15 cm. Equation (2.6) yields a value around 3×10^{-4} rad, or approximately 0.02°. This represents an extremely well collimated beam. Even at a large distance from the laser, the beam radius will remain small. For example, at a distance of 1 km, the beam would have a radius around 0.3 m.

Equation (2.6) presents the lower limit set by diffraction for a Gaussian beam. We see that the diffraction limit is directly proportional to the wavelength and inversely proportional to the aperture of the laser. If a laser has a beam divergence angle approximately equal to the minimum value defined by Equation (2.6), the laser is said to be diffraction limited.

Strictly, Equation (2.6) applies only to Gaussian beams with infinite extent. In any real laser, the Gaussian beam will be truncated, that is, the outer edges of the beam are cut off by the aperture of the laser. The effect of the truncating aperture will modify the diffraction limit slightly.

Larger values of the beam divergence angle apply if the beam profile is other than a truncated Gaussian. The beam divergence angles associated with other beam spatial profiles (see Section 2.C) are larger than the value given by Equation (2.6). In many practical cases, lasers with relatively low output power will emit diffraction-limited beams. However, higher-power lasers often emit beams that are not Gaussian and the beam divergence angle will be larger than that given by Equation (2.6). The beam divergence angle also tends to increase with increasing laser power. We will return to this point in the next section.

As a specific example, the beam divergence angle of a plane wave emerging through a circular aperture of diameter D is

$$\theta = \frac{1.22\lambda}{D} \tag{2.7}$$

This is a familiar result often quoted in optics texts. In the case where the beam waist of a Gaussian beam nearly fills the aperture, so that $2w_0 \approx D$, the functional form of Equation (2.7) becomes the same as that of Equation (2.6). Equations (2.6) and (2.7) differ only in the value of the numerical constant.

Table 2-2 **Beam Divergence Angles for Typical Lasers**

Laser Beam	He–Ne	Ar	CO_2	Ruby	Nd:glass	Nd:YAG	AlGaAs
divergence (mrad)	0.5–1.0	0.5–1.0	1–5	1–5	0.5–10	1–10	20 × 200

Table 2-2 lists typical beam divergence angles for some common lasers. In many cases, there is a spread of approximately an order of magnitude. The spread is due to construction of lasers with different diameters. For the higher-power lasers, like ruby, Nd:glass, and CO_2, the beam divergence angle will often be larger than the diffraction-limited value given by Equation (2.6). Note also that semiconductor lasers (AlGaAs) often have rather large divergence angles, with elliptical profiles.

The values in Table 2-2 are those of the raw beam emerging from the laser without further optics. One may obtain better collimation by passing the beam backward through a telescope, that is, putting the light into the eyepiece of a telescope. Schematic diagrams for this are shown in Figure 2-6. These represent the so-called Keplerian and Galilean configurations. In each case, the lenses are separated by the sum of their focal lengths, f_1 and f_2. Note that f_1 is less than zero for the diverging lens in the Galilean configuration. The Keplerian telescope is not suitable for use

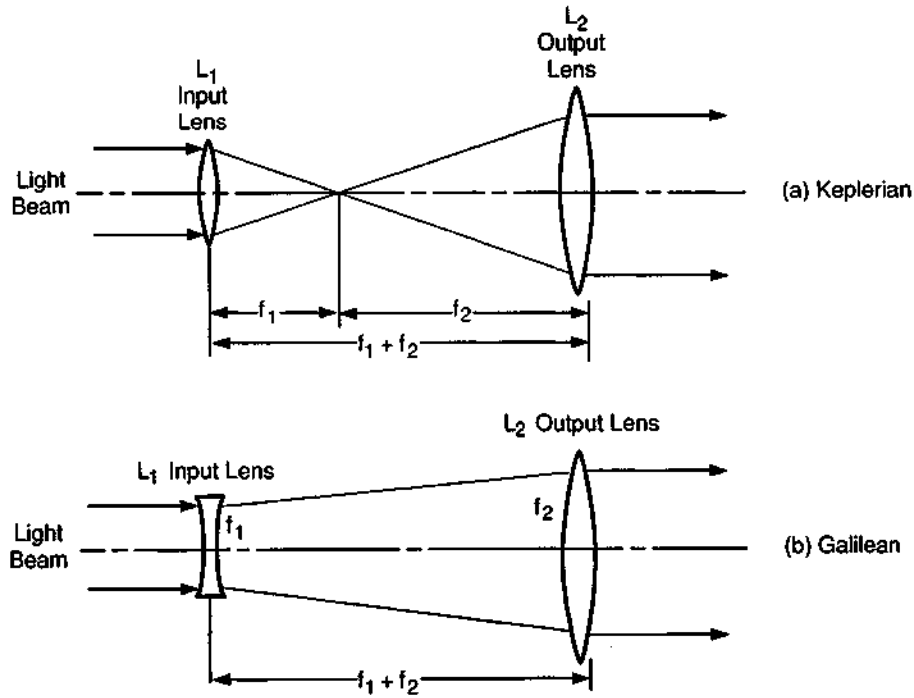


Figure 2-6 (a) Keplerian beam expander. (b) Galilean beam expander.

with high-power lasers because the beam is focused to a small spot and the high irradiance at the focus may cause gas breakdown and produce an absorbing plasma.

The beam will be collimated by a factor equal to the inverse of the magnification of the telescope. Thus, one has

$$\theta_i D_i = \theta_f D_f \tag{2.8}$$

where the subscripts i and f refer to the incident beam and emerging beams, respectively, and D and θ are respectively the beam diameter and the beam divergence angle. This equation assumes that the beam fills the aperture of the telescope. Note that the diameter of the beam increases in the same proportion that the collimation is improved. Thus, these collimating telescopes are often called beam expanding telescopes.

Let us consider a numerical example. If a diffraction-limited Gaussian helium–neon laser has a tube with a diameter of 3 mm, the raw beam diameter would be approximately 2×10^{-4} rad. If this beam is collimated by a beam expanding telescope with an output diameter of 10 cm, the beam divergence angle will be reduced to about 6×10^{-6} rad. At the distance of the moon, the beam would have expanded to a diameter around 1.5 miles.

C. Spatial Profiles of Laser Beams

There are certain distinctive spatial profiles that characterize the cross sections of laser beams. Gas lasers in particular usually have well-defined symmetry. The spatial profiles characteristic of lasers are termed transverse modes and are represented in the form TEM_{mn}, where m and n are small integers. The term TEM stands for "transverse electromagnetic." The transverse modes arise from considerations of resonance inside the laser cavity and represent configurations of the electromagnetic field determined by the boundary conditions inside the cavity. In this respect, the transverse modes are similar to the familiar case of resonant modes in a microwave cavity. Because of the large difference in wavelength, many optical wavelengths are contained within the laser cavity, whereas usually only a few wavelengths lie within a microwave cavity.

Because the laser resonator is usually elongated, with the length between the mirrors much greater than the lateral dimensions, one often considers that the field configurations are separated into transverse and longitudinal modes that are nearly independent of each other. The longitudinal modes (the field configurations along the axis) have been described in Chapter 1. In the direction normal to the laser axis, the field inside the cavity often consists of transverse configurations that propagate back and forth along the laser axis. The pattern of the transverse field configuration does not change after completion of the longitudinal round-trip between the two mirrors. These patterns are the TEM modes.

In a rectangular geometry, the mode numbers m and n represent the number of nulls in the electric field in the two dimensions perpendicular to the laser axis. The

transverse electric field distribution $E(x, y)$ in the directions x and y perpendicular to the axis are given by

$$E(x, y) = E_0 H_m\left[\left(\frac{x}{w_0}\right)2^{1/2}\right] H_n\left[\left(\frac{y}{w_0}\right)2^{1/2}\right] \exp\left[-\frac{x^2 + y^2}{{w_0}^2}\right] \tag{2.9}$$

where E_0 is a constant amplitude factor, $\mathrm{H}_n(x)$ is the Hermite polynomial of order n, and w_0 is the beam radius. The polynomial $\mathrm{H}_n(x)$ is defined by

$$H_n(x) = (-1)^n \exp(x^2)\left(\frac{d^n}{dx^n}\right)\exp(-x^2)$$

$$= \mathrm{n}! \sum_{i=0}^{[n/2]} \frac{(-1)^i (2x)^{n-2i}}{i!(n-2i)!} \tag{2.10}$$

with $[n/2] = n/2$ for n even and $[n/2] = (n-1)/2$ for n odd. Some Hermite polynomials of low order are

$$\begin{aligned} \mathrm{H}_0(x) &= 1 \\ \mathrm{H}_1(x) &= 2x \\ \mathrm{H}_2(x) &= 4x^2 - 2 \end{aligned} \tag{2.11}$$

The parameter w_0 is called the beam radius or the spot size. It measures the beam radius of the lowest-order transverse mode, for which $m = n = 0$. This is commonly called the fundamental mode, and according to the foregoing equations, the field distribution is described by a Gaussian function. The beam radius is the distance from the axis to the point at which the field amplitude is reduced by a factor $1/e$. Because the power is proportional to the square of the electric field, the distribution of the power in the beam will decrease by a factor of $1/e^2$ from its value at the center of the beam. The spatial profile for the beam intensity in this lowest-order Gaussian mode is

$$I(r) = I(0) \exp\left(\frac{-2r^2}{{w_0}^2}\right) \tag{2.12}$$

where $I(r)$ is the intensity of the beam as a function of radius r from the central axis.

Some examples of the spatial distribution of the light intensity in some rectangular transverse mode patterns are shown in Figure 2-7. The Gaussian TEM_{00} mode has no nulls in either the horizontal or vertical direction. The TEM_{10} mode has one null in the horizontal direction and none in the vertical direction. The TEM_{20} mode has two nulls in the horizontal direction and none in the vertical direction. The TEM_{11} mode has one null as one passes through the pattern either vertically or horizontally. The patterns represented in Figure 2-7 are commonly encountered in many practical lasers. In some cases, a superposition of a number of these modes can be present simultaneously; thus, the pattern of light intensity can become complicated.

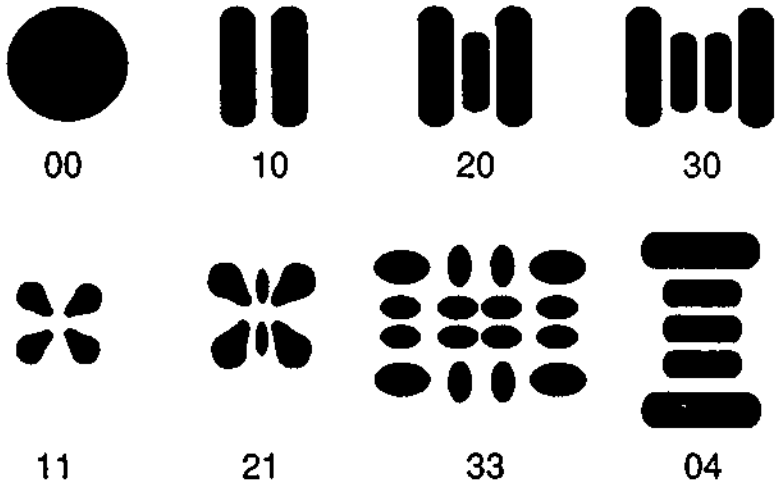


Figure 2-7 Transverse mode patterns in rectangular symmetry. The integers denote the number of nulls in light intensity as one passes through the pattern in each of two mutually perpendicular directions.

We note that the frequency of operation of the various TEM_{mn} modes changes slightly as m and n change, according to Equation (1-16). Thus, each longitudinal mode in the frequency spectrum shown in Figure 1-16, specified only by the integer q, will be split into a number of closely spaced components, specified by m, n, and q.

Now let us consider in more detail the important Gaussian mode. It is obvious that operation in the Gaussian mode is desirable because of the simple spatial profile and the lack of nulls in the distribution of the intensity across the beam.

Another fact of interest is that Gaussian beams propagate as Gaussian beams. The spatial profile of the Gaussian beam will retain its Gaussian form as the beam propagates and is transmitted through optical systems. Other higher-order modes will not retain their original spatial distribution as the beam propagates.

Also, as we discussed in Section 2.B, Gaussian beams have the minimum possible beam divergence angle, as defined by Equation (2.5). If the observed far-field beam divergence angle corresponds to this value (or perhaps exceeds it only very slightly), the beam is said to be "diffraction limited." Only Gaussian beams are truly diffraction limited. Other TEM_{mn} modes (m or $n > 0$) have larger beam divergence angles in the far field. One usually desires to have as small a beam divergence angle as possible.

In addition to the patterns with rectangular symmetry shown in Figure 2-7, there are solutions of the field equations that allow circular symmetry. In practice, lasers often operate in the patterns with rectangular symmetry, even if the cross section of the laser is round. Some examples of cylindrical TEM modes are shown in Figure 2-8. We note that the TEM_{00} Gaussian mode is identical in rectangular and cylindrical symmetry.

The mode denoted by TEM_{01}^{*} represents a superposition of two similar modes rotated by 90° about the axis relative to each other. Thus, TEM_{01}^{*}, often called the donut mode, is made up of a combination of TEM_{01} and TEM_{10} modes.

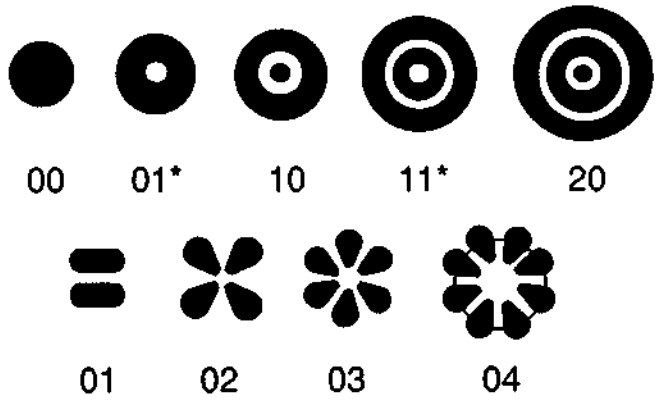


Figure 2-8 Transverse mode patterns in cylindrical symmetry. The first integer labeling the mode pattern indicates the number of nulls as one passes from the center of the pattern to the edge in a radial direction. The second integer specifies half the number of nulls as one moves in an azimuthal direction. The modes marked with an asterisk are linear superpositions of two modes rotated by 90° about the central axis.

Sometimes lasers will operate in a high-order transverse mode or perhaps in a number of modes simultaneously. Slight imperfections, such as dust particles or mirror misalignment, can cause a lack of symmetry and favor operation in high-order modes. This is undesirable because the beam divergence angle then becomes larger than the diffraction limit. Laser manufacturers now commonly design their lasers to favor operation in the TEM_{00} mode, at least for relatively low power lasers. The most common method is to limit the diameter of the laser medium. The diameter is kept small enough so that the beam waist of the TEM_{00} mode is entirely contained within the laser medium, but the higher-order TEM_{mn} modes are not completely contained. This introduces a factor that discriminates against operation of the higher-order modes.

If a laser does in fact operate in a high-order mode, it sometimes can be made to operate in the Gaussian TEM_{00} mode by inserting a circular aperture within the cavity. Apertures of progressively smaller size may be inserted until one reaches a diameter where the laser operates only in the TEM_{00} mode. This is generally accomplished at some expense in output power, but the accompanying decrease in beam divergence angle and simplicity of the spatial pattern may be adequate compensation for this decrease.

A quantitative measure of beam quality that is frequently employed is called M^2. This parameter compares the divergence of the given beam with that of a pure Gaussian beam with the same waist located at the same position. For a pure Gaussian beam, the value of M^2 is equal to 1. If there is an admixture of higher-order modes present, the value of M^2 will be greater than 1. Beams that are close to Gaussian will have a value of M^2 only slightly greater than unity. Beams with poorer quality may have a value of M^2 several times unity.

The value of M^2 may be measured by determining the location of the beam waist and its size, and making measurements of the beam diameter at points along the

direction of propagation. Commercial instrumentation is available that can provide automatic determination of the value of M^2 for a given input beam.

We note explicitly that the word "mode" has two meanings. The spatial patterns shown in Figures 2-7 and 2-8 are the transverse modes of the laser. The cavity modes referred to in Figure 2-1 are longitudinal (or axial) modes. It is possible for the laser to operate in several different transverse modes and several longitudinal modes at the same time. A laser also may operate in a single transverse mode and several longitudinal modes, or in a single longitudinal mode and several transverse modes.

In Chapter 1 we discussed the unstable resonators that are sometimes used with high-power gas lasers. For a long slim laser, with small Fresnel number N_F, a single low-order mode can extract all the power. But if N_F is much greater than unity, a larger number of high-order modes can be present. The lowest-order mode will extract only about $1/N_F$ of the available power. This means that a number of high-order modes will be present, and the beam divergence angle will be large. For a large Fresnel number, the unstable resonator is better suited to output coupling in a low-order mode with small divergence angle.

Figure 2-9 shows the integrated output power as a function of the far-field beam divergence angle from a CO_2 laser experiment [3] with an unstable resonator. This shows the far-field pattern and indicates a well-collimated beam. The experimental

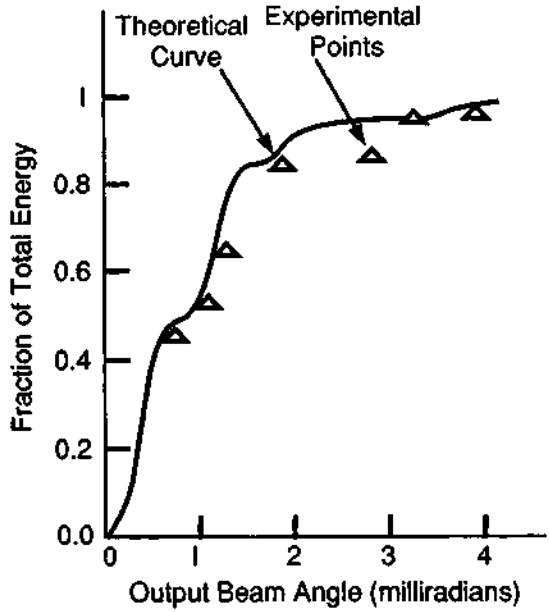


Figure 2-9 Fraction of total beam energy contained within the indicated beam angle for a CO_2 laser operating with an unstable resonator. (From J. Davit and C. Charles, *Appl. Phys. Lett.* 22, 248 (1973).)

points are in reasonably good accord with the theory that has been developed for unstable resonators.

This description of TEM modes is relevant only for lasers operating at relatively low output power. Increases in output power are usually achieved at the expense of a more complex spatial structure. This is illustrated in Figure 2-10, which shows how the transverse mode structure changes in a relatively small continuous Nd:YAG laser as the power increases. At low power, the laser operates in the TEM_{00} mode. At higher values of power, the laser shifts to progressively higher-order transverse modes. This is accompanied by a progressive increase in the beam divergence angle.

In some cases the spatial profile may be even more complicated, and not describable in reasonably simple mathematical terms like the TEM modes. High-power solid state lasers especially, like ruby and Nd:glass, can exhibit very complicated patterns. These patterns are the result of inevitable imperfections in the laser medium, such as inhomogeneities in the index of refraction of the laser rod. An example of such a complicated pattern observed in the operation of a Q-switched ruby laser is shown in Figure 2-11. It is apparent that there are "hot spots" where the irradiance is much higher than the average value. It is a fact of life that very high power in solid state lasers is accompanied by complicated beam profiles.

Single-transverse-mode operation may be obtained for solid state lasers by limiting the power and by inserting apertures within the laser. It is also important to

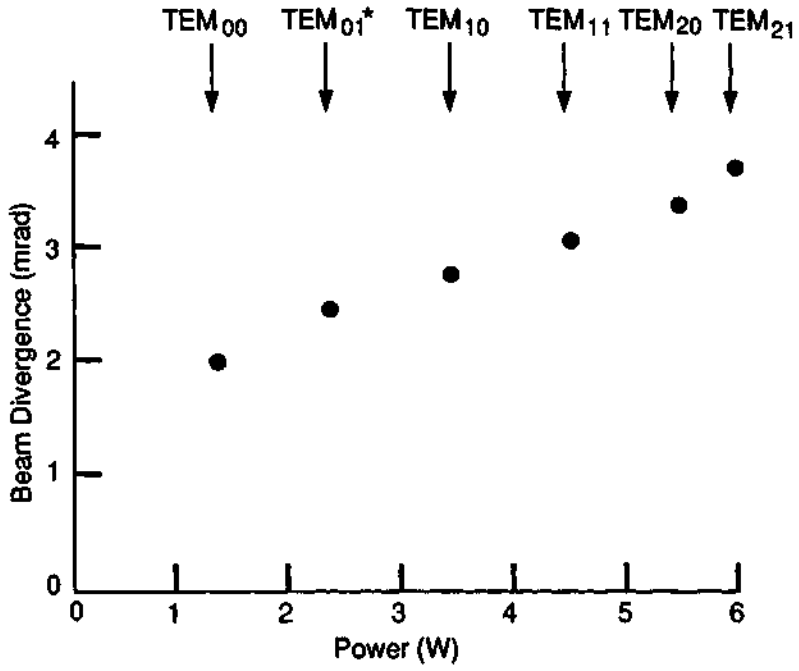


Figure 2-10 Variation in transverse mode structure and beam divergence angle with output power for a commercial continuous Nd:YAG laser.

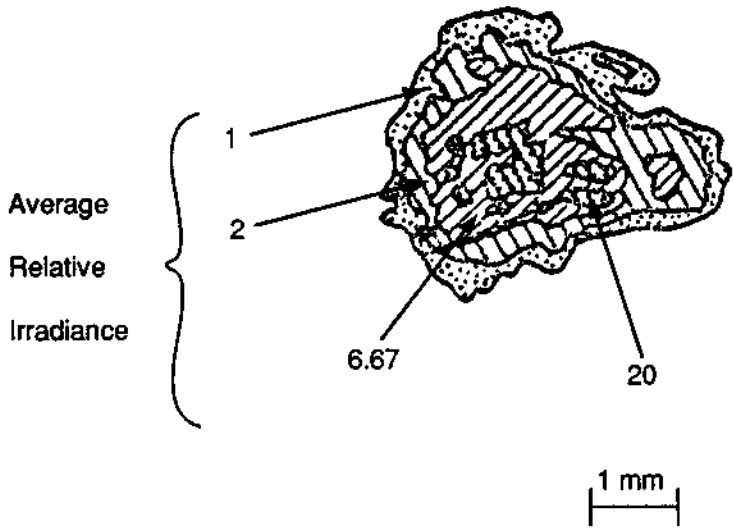


Figure 2-11 Contour of irradiance in the beam from a Q-switched ruby laser. Contours of equal average irradiance are indicated in arbitrary units.

choose laser rods with high optical quality and homogeneous index of refraction. As an example [4], a ruby laser of common design was controlled so as to operate in the TEM_{00} mode by insertion of a 2-mm-diameter aperture, yielding pulses with 20 nsec duration and 2 MW peak power.

The spatial patterns from semiconductor lasers are of particular interest. The light emerges typically in a fan-shaped beam, with the broader dimension of the fan perpendicular to the narrow dimension of the junction. This pattern is the result of diffraction by the narrow slit defined by the junction in the semiconductor. This will be described in more detail in Chapter 3.

D. Temporal Behavior of Laser Output

There are a number of distinctive types of temporal behavior exhibited by lasers. An understanding of the different time sequences that commonly occur in the outputs of various lasers is important. When one chooses a particular laser for an application, the availability of a laser with a particular range of pulse duration may be a significant factor.

Many types of lasers can operate continuously. In commonly used parlance, these devices are often called CW (continuous wave). Continuous operation is available for most lasers using a gas as the active medium. In fact, the most common method of operation for most gas lasers is CW. In addition, some solid state lasers, like Nd:YAG, some dye lasers, and many semiconductor lasers may be operated continuously. Some solid state lasers, like ruby and Nd:glass, are rarely continuous. For example, small experimental ruby lasers have been operated continuously, but these devices were laboratory curiosities.

Even during continuous operation, there may be fluctuations in the amplitude of the output power. The fluctuations may have a frequency spectrum extending out to hundreds of kilohertz, or perhaps even megahertz. These fluctuations may arise from thermal effects or from different longitudinal modes going in and out of oscillation. The argon laser, in particular, is subject to fluctuations in its output unless it is thermally stabilized. Typical modern He–Ne lasers after warm-up may have peak-to-peak fluctuations of a few tenths of one percent of the output power, whereas argon lasers may have fluctuations of one percent or more. The variations also may be greater in older models of continuous lasers.

Gas lasers may be operated pulsed under some circumstances. Indeed, for the nitrogen laser and the excimer lasers, the only type of operation available is pulsed operation. CO_2 lasers are often operated pulsed in order to extract high levels of output power.

For solid state lasers (like Nd:glass), which are often pumped by flashlamps, there are several distinctive regimes of pulse operation, which are described in the following subsections.

1. NORMAL PULSE OPERATION

In the so-called normal pulse mode of operation for solid state lasers, no control is exerted on the pulse duration, other than by control of the duration of the flashlamp pulse. The light is allowed to emerge from the laser at its own natural rate, without use of devices that change the properties of the resonator. In this type of operation, typical pulse durations are in the range from 100 μsec to several milliseconds. The pulse duration is controlled primarily by changing the inductance and capacitance of the power supply for the flashlamp, so as to change the duration of the pumping pulse. This type of operation is also sometimes called long-pulse operation or free-running operation.

In many cases, normal pulse solid state lasers exhibit substructure within the pulse. The substructure consists of spikes with durations typically in the microsecond regime. These spikes are often called relaxation oscillations. The spacing and amplitude of the relaxation oscillations are usually not uniform. An example of this behavior is shown in Figure 2-12 [5]. The upper trace shows the pulse envelope, which lasts several hundred microseconds. These high-frequency spikes are superimposed on the general pulse envelope. The lower trace shows a portion of the pulse on an expanded time scale to show the behavior of the individual microsecond duration pulsations in the train of relaxation oscillations. Most free-running solid state lasers exhibit this type of behavior to some extent. In some cases, the pulsations may be evenly spaced, and sometimes they may be present near the beginning of the pulse and damp out near the end. Sometimes, they may be present only as a ripple superimposed on the millisecond duration pulse; often, they may represent a complete modulation with the minimum amplitude going to zero. But they are a persistent feature of free-running solid state laser operation.

In some applications, the presence of the relaxation oscillations may be desirable. The peak power reaches values much higher than the average of the pulse envelope and can initiate surface disruption more easily in material processing. For other

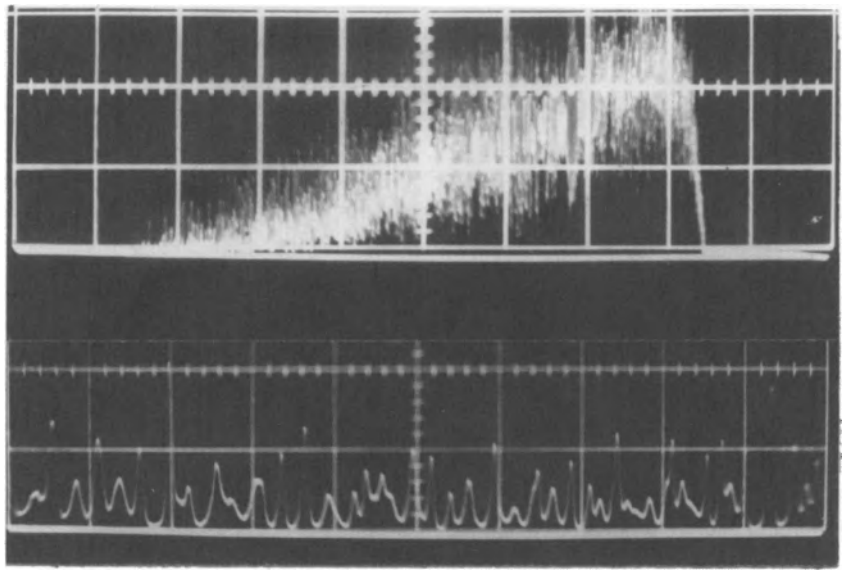


Figure 2-12 Relaxation oscillation behavior in a pulsed ruby laser. The upper trace shows the entire pulse at a sweep speed of 200 μsec/division. The lower trace shows an expanded portion of the pulse at a sweep speed of 10 μsec/division. Time increases from right to left. (From J. F. Ready, *Effects of High Power Laser Radiation,* Academic Press, New York, 1971.)

applications, particularly when good control of the pulse characteristics is desired, the presence of the relaxation oscillations is undesirable and the laser designer may attempt to suppress them.

What may be considered to be normal pulse operation for gas lasers consists primarily of turning the power supply on and off rapidly. Typical pulse durations are in the microsecond to millisecond regime. It is difficult to modulate the high-voltage power supplies more rapidly. High-power gas lasers that are pulsed by means of the power supply often will exhibit an initial spike with peak power perhaps two or three times the value of the average power during the pulse. This is illustrated in Figure 2-13. The value of the power after this initial spike is typically what one would obtain during continuous operation. Again, in material processing applications, the initial spike may be useful in initiating surface breakdown.

2. *Q-SWITCHED OPERATION*

A factor called Q is often used to describe the properties of the resonant cavity. This factor characterizes the ability of the cavity to store radiant energy. A large value of Q means that energy is stored well inside the cavity. A low value means that energy in the cavity will emerge rapidly. For example, if the mirrors on the ends of the cavity have high reflectivity, Q will be relatively high; if the mirrors have lower reflectivity, whatever energy is present will emerge rapidly, and Q will be low.

Let us consider a situation in which Q is high initially during the time that the active medium is being excited and then is suddenly reduced to a low value. In this

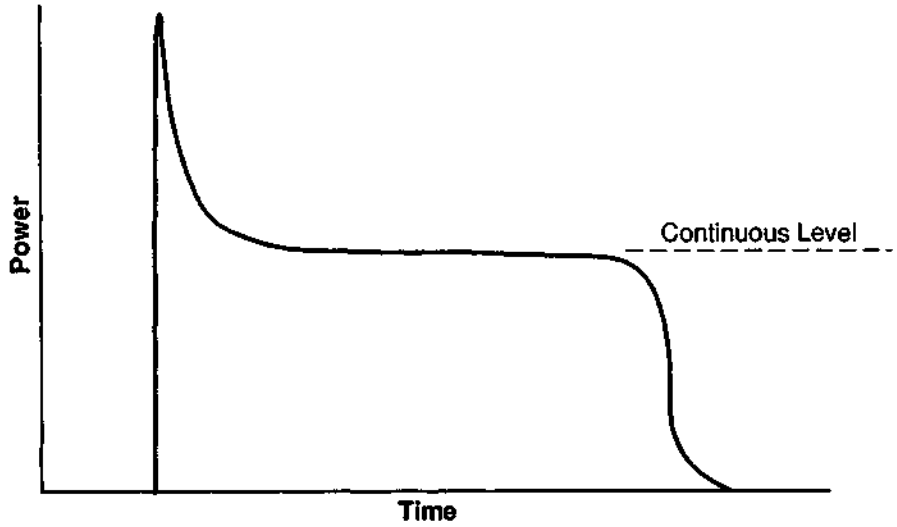


Figure 2-13 Typical behavior of the output from a gas laser pulsed by rapidly pulsing the voltage on the power supply. The initial spike is shown.

situation, a large amount of energy will be stored within the cavity, during the time when Q is high. Very little will emerge through the mirrors. Then, when Q drops to a low value, the stored energy will rapidly be transmitted through the mirrors. Because a large amount of energy emerges in a short period of time, the power output will be high. Such lasers are called Q-switched lasers. (Sometimes the synonymous terms Q-spoiled laser or giant pulse laser are used.) For a laser cavity consisting of a 100 percent reflecting mirror and an output mirror with reflectivity R, the factor Q is given by

$$Q = \frac{2D\omega}{(1-R)c} \tag{2.13}$$

where D is the distance between the mirrors, ω is the angular frequency of the light, and c is the velocity of light. According to this equation, when R is high, close to unity, Q will be high. When R is appreciably less than unity, Q will be lower. Q-switching methods usually switch the effective reflectivity of one of the mirrors from a high to a low value, at a time when the laser medium has been pumped to a highly excited state.

A variety of practical techniques have been devised to change the Q of a laser cavity. They include use of rotating mirrors, electrooptic elements, acoustooptic switches, and bleachable dyes. In this section, we describe the use of a rotating mirror in order to illustrate the basic concept. A more complete description of the various Q-switches will be given in Chapter 5.

Figure 2-14 shows schematically how one uses a prism as a rotating mirror for a Q-switch. A totally reflecting prism is used as the high-reflectivity mirror. It spins about an axis of rotation that lies in the plane of the figure. A right-angle prism (also

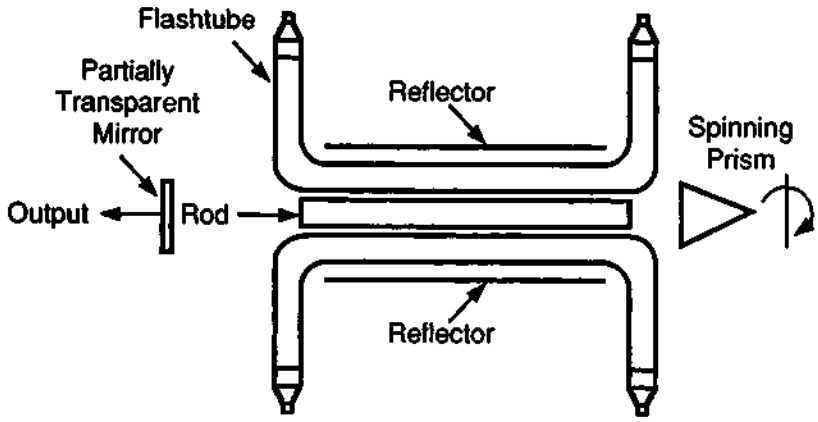


Figure 2-14 A solid state laser Q-switched with a rotating prism.

called a Porro prism) may be used for this application. The right-angle prism is retroreflecting for light that enters the prism through the face of the hypotenuse of the right triangle. Thus, the prism has the property that a small misalignment about the axis perpendicular to the plane of the figure does not affect the reflection. Light is reflected accurately back along the direction from which it came, even if there is some misalignment. Thus, the operation of the laser is not affected by vibration.

The reflection is highly sensitive to orientation about an axis parallel to the hypotenuse and in the plane of the figure. Thus, as the prism spins about this axis, as shown in the figure, the laser cavity is not aligned, and light that builds up in stimulated emission does not reach the output mirror. There is no resonant cavity, and laser operation cannot occur. Absorption of light from the flashlamps excites the laser rod, and the population inversion builds up to a high value, higher than that given by the normal threshold condition. Then, in the course of its rotation, the prism swings momentarily through a position where the hypotenuse of the prism is parallel to the output mirror, and the resonant cavity is established briefly. The laser rod, in its highly excited state, has a resonant cavity, and laser operation builds up rapidly. The energy stored in the rod is emitted is a short pulse with higher peak power and shorter duration than if Q-switching had not been used.

The Q-switched pulse yields much higher peak power than the longer normal pulse, but it involves extraction of a smaller amount of energy. Q-switching usually involves some reduction in total output energy, but the pulse duration is shortened considerably. Thus, the peak power becomes much higher. Table 2-3 compares the energy, power, and pulse duration available from a typical small ruby laser. These are not the highest values achieved; rather, they represent easily attainable values. The numbers in the table represent the output of the same laser with and without a Q-switching device.

Q-switching has been employed with most types of lasers, but it is most common with the high-power solid state lasers, like Nd:glass, Nd:YAG, and ruby.

Table 2-3 **Typical Values for Ruby Lasers**

Operation	Pulse energy (J)	Pulse duration (sec)	Peak power (W)
Normal pulse	10	500×10^{-6}	2×10^{4}
Q-switched pulse	1	50×10^{-9}	2×10^{7}

Q-switching has been used with CO_2 lasers, but its use with other types of gas lasers or with semiconductor lasers is uncommon.

3. *MODE-LOCKED OPERATION*

There may be additional substructure within Q-switched laser pulses. This arises because of a phenomenon called mode-locking. If there are a number of longitudinal modes present in the laser output, these modes may interfere and lead to an oscillatory behavior for the output. Thus, within the envelope of the Q-switched pulse, there may be a train of shorter pulses. This is illustrated in Figure 2-15, in which part (a) shows the smooth envelope of a Q-switched pulse, without mode-locking. If only a single longitudinal mode is present in the laser output, there will be no substructure to the Q-switched pulse, and the pulse will appear as shown in (a).

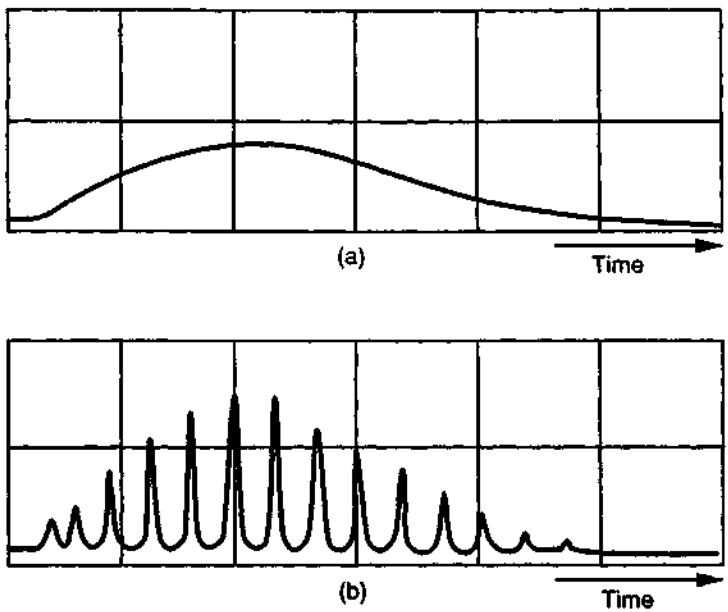


Figure 2-15 (a) Q-switched pulse without mode-locking. (b) Q-switched pulse with mode-locking.

In many lasers, a number of longitudinal modes will be present simultaneously. Ordinarily, the phases of these modes will be independent of each other. But under some conditions, the modes can interact with each other and their phases can become locked together. Then the behavior of the output will be as shown in part (b) of Figure 2-15, with a train of short, closely spaced pulses contained within the Q-switched pulse.

Mode-locking may be accomplished intentionally by introducing a variable loss into the cavity, such as an electrooptic modulator (see Chapter 5) so that the gain of the cavity is modulated at the frequency $c/2D$, which is the difference in frequency between adjacent longitudinal modes. Under many conditions, mode-locking occurs spontaneously, so that the behavior shown in Figure 2-15b is frequently encountered. Mode-locking in a given laser may be erratic, with one Q-switched pulse having mode-locking and another pulse not exhibiting it. Conditions for reliable production of mode-locking have been identified, so that reproducible mode-locking may be obtained.

In the time domain, the output of the mode-locked laser consists of a series of very short pulses. From a simple point of view, one may consider that there is a very short pulse of light bouncing back and forth between the laser mirrors. Each time the pulse reaches the output mirror, a mode-locked pulse emerges from the laser. The result is a series of short pulses separated in time by $2D/c$, the round-trip transit time of the cavity. The duration of these pulses is very short, often of the order of a few picoseconds. Hence, a laser operating in this fashion is often called a picosecond pulse laser.

One may obtain a single pulse of picosecond duration by using an output coupler such as an electrooptic switch that opens long enough to let one pulse emerge and then closes again. This pulse may be further amplified by passing it through an optical amplifier, which basically consists of an excited laser rod without its resonant cavity. In this way, picosecond pulses with very high peak power (up to terawatts) have been obtained. Such devices have been used for research in laser-assisted thermonuclear fusion, but they have found little use in industrial applications.

The minimum duration for a mode-locked pulse is equal to $1/\Delta\nu$, where $\Delta\nu$ is the spectral width of the active medium. For a Nd:glass laser, the spectral width is large, around 10^{12} Hz. The minimum pulse duration for such a laser is then around 1 psec.

The shortest pulses that have been produced have come from liquid dye lasers, for which the spectral width can be very large. Pulses as short as 30 femtoseconds (fsec) (30×10^{-15} sec) have been produced. This is remarkable in that the total duration of the pulse contains only a few cycles of the high-frequency electric field.

4. CAVITY DUMPING

In addition to Q-switching, there is another method for producing short pulses with duration in the nanosecond to microsecond regime. In this method, called cavity dumping, the laser is excited continuously. The resonant cavity has high Q, so that little energy emerges and the laser light simply remains in the cavity. An electrooptic or acoustooptic device periodically switches pulses out of the cavity. The advantage

of cavity dumping as compared with Q-switching is that it allows pulsing at higher pulse repetition rates. Cavity dumping has often been employed with Nd:YAG lasers to yield pulses at rates up to the megahertz regime. At high repetition rates, the average power of the cavity-dumped laser may be very close to the continuous power that could be obtained from the same laser, but the peak power may reach a value much higher than the average power.

5. *SUMMARY*

We have discussed a variety of different types of pulsed operation. Solid state lasers, like Nd:glass or ruby, are most often operated pulsed. In many cases the pulse repetition rate will be low, one pulse per second or less. For very high peak power lasers, there may be long intervals between isolated pulses. The Nd:YAG laser may be operated in a repetitively Q-switched mode, at fairly high pulse repetition rates, up to kilohertz.

Gas lasers, like He–Ne and argon, are usually operated continuously, but under some conditions they could be pulsed. CO_2 lasers have been operated both pulsed and continuously. Gas lasers are conveniently pulsed simply by pulsing the power supply.

To summarize the situation, Table 2-4 shows some commonly encountered methods of operation for a variety of different lasers. Typical values for pulse duration are also given. Pulse durations for common devices may extend over a range covering many orders of magnitude.

E. Coherence

Coherence is another unusual property of laser light that influences possible applications. The concept of coherence is linked to the orderliness of the light waves. An

Table 2-4 **Pulse Properties of Common Lasers**

Laser	Mode of operation	Typical pulse duration (sec)
Ruby	Normal pulse	10^{-4}–10^{-3}
	Q-switched	3–5×10^{-8}
Nd:glass	Normal pulse	10^{-3}–10^{-2}
	Q-switched	5×10^{-8}
	Mode-locked	10^{-12}–10^{-11}
Nd:YAG	Continuous	—
	Q-switched	10^{-8}–5×10^{-7}
Argon	Continuous	—
CO_2	Continuous	—
	Normal pulse	10^{-4}–10^{-2}
Liquid dye	Continuous	—
	Q-switched	10^{-9}
	Mode-locked	10^{-14}–10^{-12}

analogy sometimes used is that of ripples spreading from a stone thrown into a pond of water. The waves that spread out from the spot where a single stone enters the water move in a regular circular pattern. This is analogous to coherence. But the waves from a handful of stones are irregular and confused. This is analogous to incoherence.

Actually, the concept of coherence is more complicated than this. Much mathematical development has gone into formulating the description of coherence. It is not appropriate to review this entire body of mathematics here. We simply state the result that coherence can by specified quantitatively in terms of a function called the mutual coherence function, denoted $\gamma_{12}(\tau)$. This quantity, which is a complex number, measures the correlation between the light wave at two points, P_1 and P_2, at different times, t and $t + \tau$. Thus, the mutual coherence function has both spatial and temporal contributions.

The absolute value of $\gamma_{12}(\tau)$ lies between 0 and 1. If the value is 1, the light is completely coherent, and there is perfect correlation between the light waves at the two different points and times. The value 0 corresponds to complete incoherence, that is, complete lack of any correlation. In practice, neither limiting value is achieved, but only approached. Thus, in practice one always has partial coherence, that is, $0 < |\gamma_{12}(\tau)| < 1$. For single transverse modes in gas lasers, $|\gamma_{12}(\tau)|$ can be very close to unity, so that the light is close to being completely coherent.

The value of $|\gamma_{12}(\tau)|$ can most easily be found by the evaluation of fringe patterns formed by interference of two beams of light that have traveled two different paths. The visibility V of the fringe pattern is defined by

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{2.14}$$

where $I_{\max}$ is the maximum intensity of the light at a bright fringe and $I_{\min}$ is the minimum intensity in a dark area. If the two waves have equal intensity, the visibility is equal to $|\gamma_{12}(\tau)|$. Thus, measurement of intensity in an interference pattern can provide a straightforward method of determining the degree of coherence of laser light.

Coherence is important in any application in which the laser beam is split into parts. Such applications are common. They include distance measurements, in which the light is split into two beams that traverse paths of unequal distance, and holography, in which the beams traverse paths that may have approximately equal lengths but different spatial contributions.

If the coherence is less than perfect, the fringe patterns formed in the interferometer or hologram will be degraded. Spatial coherence also influences how wide an area of the wavefront may be used for making a hologram. Thus, coherence is of practical importance for many laser applications.

The function $\gamma_{12}(\tau)$ has both spatial and temporal dependence. If one splits a beam of light into two parts, allows the beams to traverse paths of different length, and then recombines the beams, one is dealing with temporal coherence, The time delay τ equals $\delta L/c$, where δL is the difference in path length and c is the velocity of light. Thus, one is dealing with the function $\gamma_{11}(\tau)$, that is, the coherence in the light

wave at one point in the wave but at times separated by τ. This gives the correlation between the light wave at one instant of time and one point in space and the light wave at a later time at the same point.

The phase of the optical wave plays an important role in coherence. The light wave at a point given by the position vector $\mathbf{r}$ may be represented by the equation

$$E(\mathbf{r}, t) = E_0(\mathbf{r}, t) \cos[\omega t + \phi(\mathbf{r}, t)] \tag{2.15}$$

where E_0 is the amplitude, ϕ a phase function, and ω the angular frequency. The function E_0 is relatively constant. Most of the fluctuations in the beam are phase fluctuations. If the phase function ϕ changes randomly, there will be imperfect correlation between the light wave at one instant and the same light wave at a later instant. Then $|\gamma_{11}(\tau)|$ will be reduced from unity.

The time period Δ in which the phase undergoes random changes is called the coherence time, and it is related to the linewidth $\Delta\nu$ of the laser by

$$\Delta t \approx 1/\Delta\nu \tag{2.16}$$

If measurements are made in a time short compared with the coherence time, the value of $|\gamma_{11}(\tau)|$ will be high. This is also true if the path difference between two interfering beams is less than $c\,\Delta t$, a quantity that is called the coherence length. The coherence length is a measure of the propagation distance over which the beam stays coherent. For measurements that extend over longer times (or over longer path differences), $|\gamma_{11}(\tau)|$ will be reduced, approaching zero for times very long compared with the coherence time.

The presence of multiple longitudinal modes in the output of a laser broadens the spectral width and reduces the coherence length. This in turn reduces fringe visibility in a holographic or interferometric experiment. The influence of multimode oscillation is shown in Figure 2-16, which shows fringe visibility as a function of path difference for a 1-m-long laser operating in different numbers of longitudinal modes. This figure is obtained from the equation [6]

$$V = \left| \frac{\sin\left(\dfrac{N\pi\,\Delta L}{2D}\right)}{N \sin\left(\dfrac{\pi\,\Delta L}{2D}\right)} \right| \tag{2.17}$$

where N is the number of longitudinal modes, ΔL is the difference in path length for the two interfering beams, and D is the length of the laser cavity.

Figure 2-16 expresses the practical result that if one uses a laser with multiple longitudinal modes, the allowable path difference between two interfering beams is limited to a few centimeters at most. In holography, the permissible subject depth would be limited. For interferometry, the distance over which a measurement could be made is restricted. Single-frequency operation of the laser would remove this limitation, as the figure also shows.

The concept of spatial coherence arises when one considers the function $\gamma_{12}(0)$, which is the function describing the coherence when light waves from two different

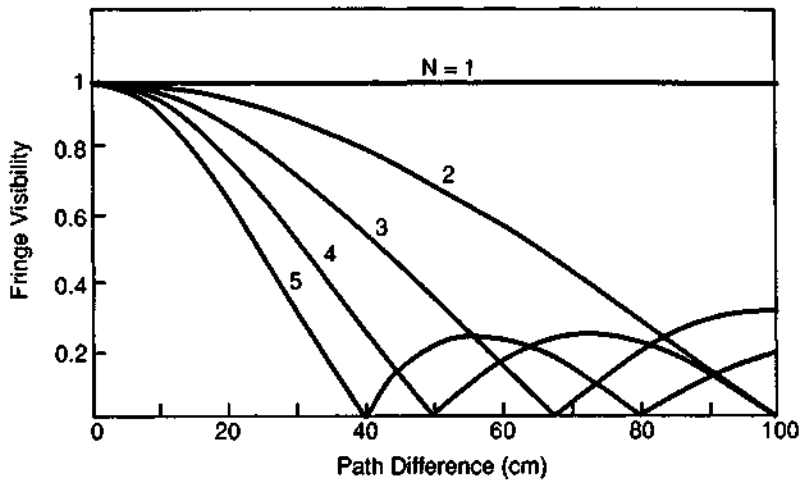


Figure 2-16 Fringe visibility in an interferometric experiment as a function of path difference for a 1-m-long laser operating in N longitudinal modes.

points on the wavefront are combined with no path difference. The properties of spatial coherence depend strongly on the transverse mode structure of the laser. The absolute value of the mutual coherence function will be very close to unity in a laser beam that consists of a single transverse mode, as is commonly the case for a well-stabilized gas laser. This implies that the beam has excellent spatial coherence over its entire cross section. But if the laser has a number of transverse modes operating simultaneously, the function $|\gamma_{12}(\tau)|$ can have any value between zero and unity, depending on which points on the wavefront are examined.

For continuous gas lasers, the coherence time is approximately equal to the reciprocal of the linewidth and the spatial coherence is close to unity for operation in a single transverse mode. Helium–neon lasers in particular usually have good coherence. A typical 1-m-long He–Ne laser operating in several longitudinal modes may exhibit a coherence time around 4×10^{-10} sec, equivalent to a coherence length $c\,\Delta t \approx 12$ cm. The coherence length is relatively small because the presence of several longitudinal modes broadens the spectral width. For a stabilized He–Ne laser operating in a single longitudinal mode, the coherence length may be much longer, many meters in some cases.

In one measurement involving two-slit interference [7], $|\gamma_{12}(0)|$ was in excess of 0.9985, for a continuous He–Ne laser operating in the Gaussian mode. When the laser was operated in the TEM_{01} mode, the value of $|\gamma_{12}(0)|$ was approximately 0.98. The decrease from unity was probably due to a weak admixture of higher-order modes. When a He–Ne laser is operated in multiple transverse modes, the value of $|\gamma_{12}(\tau)|$, as measured in a two-pinhole interference experiment, depends on the position of the pinholes in the cross section of the beam.

For pulsed lasers, the situation is more complex. In one experiment [8], spatial coherence of a pulsed ruby laser was demonstrated over a distance of around 0.005

cm. The laser operated so that only small regions ("filaments") in the ruby contributed to the emission. The result corresponded to spatial coherence over the dimension of one filament. In another study of the spatial coherence of a ruby laser [9], the values of $|\gamma_{12}(0)|$ fell in the range 0.2–0.6. The exact value depended on the type of resonator. The higher values were obtained with resonators that encouraged operation in a small number of transverse modes.

The temporal coherence of pulsed lasers may be limited by shifts in the frequency during the pulse. In an experiment [10] in which a pulsed ruby laser illuminated two slits with a variable path difference between the light beams reaching the two slits, the visibility of the fringes decreased for path differences greater than 2500 cm, and it disappeared around 3000 cm. This corresponds to a coherence time around 0.1 μsec, whereas the duration of the individual pulses was around 0.56 μsec.

In summary, continuous lasers operating in a single longitudinal and transverse mode can offer nearly ideal coherence properties. Coherence may be degraded by the presence of multiple modes or by variations introduced in pulsed operation. Table 2-5 summarizes some of these results.

A noteworthy result of the coherence of laser light is the appearance of a speckle pattern when a visible laser illuminates a diffusely reflecting surface, like a piece of

Table 2-5 **Summary of Coherence Properties**

Laser	Spatial coherence function, $\lvert\gamma_{12}(0)\rvert$	Coherence time, Δt	Coherence length, $c\,\Delta t$
Continuous, single transverse mode	~1	1/linewidth	~10 cm for $N > 1$; meters for $N = 1$ (N = no. of modes)
Continuous, >1 transverse mode	Depends on points sampled	1/linewidth	~10 cm for $N > 1$; meters for $N = 1$
Pulsed	~0.5	≤ pulse length	$\leq c \times$ pulse length

Typical Coherence Lengths

Laser	Coherence length (cm)
He–Ne (multimode)	20
He–Ne (single frequency)	100,000
Argon (multimode)	2
Argon (single frequency)	5,000
Nd:YAG	1
Nd:glass	0.02

white paper. The beam has a speckled, granular appearance. This occurs because a random interference pattern is formed from the contributions of light reflected from various areas of the surface. On a diffusely reflecting surface, there will be small hills and valleys, with dimensions of the order of the wavelength of light. The coherent light scattered from the surface will produce an interference pattern. The interference leads to bright spots where there is constructive interference and to dark spots where there is destructive interference.

The observation of the speckle pattern came as a surprise when continuous visible lasers first became available. It has the practical implication that the irregular pattern makes visible laser light a poor choice for applications involving illumination of surfaces.

Speckle patterns could in principle be produced with conventional optical sources. But it is only the laser, with its high degree of spatial and temporal coherence, that makes speckle patterns familiar.

F. Radiance

One of the significant attributes of lasers is their high radiance. The radiance of a light source is the power emitted per unit area per unit solid angle. The relevant solid angle is that defined by the cone into which the beam spreads. Lasers emit their light into very small solid angles, perhaps 10^{-6} steradians (sr) or less, whereas conventional light sources emit into a solid angle of 4π sr. Because lasers can emit high levels of power into very small solid angles, they represent sources of very high radiance. Sometimes the term brightness is used in place of radiance.

Radiance is an important quantity. A fundamental theorem of optics states that the radiance of a source may not be increased by any optical system; at best, the radiance remains equal to the radiance of the source. Moreover, high radiance is important for delivering high irradiance (power per unit area) to a workpiece. Thus, the high radiance of lasers is an important factor in their ability to perform applications like material processing.

The radiance of a laser may be limited by the presence of additional TEM_{mn} modes other than the Gaussian mode. Often, as laser power is increased, the number of modes increases (e.g., see Figure 2-10) and the radiance remains fixed or increases only slightly. The quest for high irradiance at a workpiece can thus involve improvement of mode characteristics and decreasing beam divergence angle just as much as increasing power.

The development of laser systems specifically designed for high radiance has led to diffraction-limited Nd:glass lasers consisting of a single-mode oscillator followed by amplifiers, yielding radiance as high as 2×10^{17} W/cm^2-sr. More usual values are as follows: for a He–Ne laser, 10^6 W/cm^2-sr, and for a Q-switched ruby laser, 10^{11}–10^{12} W/cm^2-sr. For comparison, the sun has radiance around 130 W/cm^2-sr.

It is appropriate at this point to summarize some of the radiometric quantities that will be used repeatedly throughout the text. These are the quantities that are useful in defining the transfer of radiant energy. Table 2-6 lists some of the important phys-

Table 2-6 **Radiometric Nomenclature**

Quantity	Standard Radiometric term*	Popularly used term	Unit
Radiant energy	Energy	—	Joule
Rate of flow of radiant energy	Flux	Power	Watt
Areal density of radiant energy incident on a surface	—	Fluence	Joule/m^2
Radiant Intensity (flux per unit solid angle)	Intensity	—	Watt/sr
Areal density of radiant intensity	Radiance	Brightness	Watt/m^2-sr
Areal density of flux incident on a surface	Irradiance	—	Watt/m^2

*Americal National Standard Nomenclature and Definitions for Illuminating Engineers, ANSI Z7.1-1967.

ical quantities and their names, including both the names recommended by the American National Standards Institute (ANSI) and some commonly encountered alternatives. The table also lists the units recommended for each quantity.

In this book we will use terms and units that are commonly employed in the laser community, even though they depart from the recommended nomenclature. Thus, we will use the term power, instead of flux, for the rate of flow of radiant energy, and the units W/cm^2 for irradiance.

G. Focusing Properties of Laser Radiation

To determine the irradiance that can be produced by laser radiation, we must consider the size of the spot to which the beam can be focused. It is not possible to focus the beam to an infinitesimal point. There is always a minimum spot size, which is determined ultimately by diffraction. Often, of course, because of imperfections in the optics, one cannot reach the limit set by diffraction, so that the spot size is larger than the following considerations indicate. But in any optical system, there is an ultimate limit, termed the diffraction limit, which determines the minimum focal area and, hence, the maximum irradiance that can be attained.

An important property of Gaussian laser beams is that they propagate as Gaussian beams; that is, they have the same intensity distribution in the near and far fields and after passage through an optical system. Gaussian beams are also uniphase; they have the same phase across the entire wavefront. A Gaussian beam can be focused to the minimum spot size, of the order of the wavelength of light. This property of uniphase Gaussian beams is very useful, and it distinguishes laser beams from incoherent light beams. Uniphase Gaussian beams can be focused to smaller spots than incoherent beams.

If the original spreading of the beam was determined by diffraction at the aperture of the laser, and if the distance from the laser to the lens is small, so that the

beam diameter has not increased much, then the minimum spot radius r_s that can be obtained when the beam is focused by a lens with focal length F is

$$r_s = F\theta \tag{2.18}$$

where θ is the beam divergence angle. This equation neglects any effects of lens aberrations. It is commonly used as a convenient approximation for estimating minimum focal size.

Because the diffraction-limited beam divergence angle is approximately $\theta = \lambda/D$, where D is the diameter of the limiting aperture, one has (assuming that the beam fills the lens aperture)

$$r_s = \frac{F\lambda}{D} = \lambda \times F\# \tag{2.19}$$

where $F\#$ denotes the F-number of the lens. Because it is impractical to work with F-numbers much less than unity, the minimum value of the spot size will be of the same order of magnitude as λ. This is the origin of the statement that laser light can be focused to a spot the size of its wavelength.

Consider a numerical example that shows how focusing properties are important in the interaction of laser radiation with material. If a 10-W gas laser with a Gaussian beam and a divergence angle of 10^{-3} rad is focused by an F:1 lens, the resulting spot would have an area around 10^{-6} cm^2, according to the preceding discussion, and the irradiance near the center of the spot would be around 10^7 W/cm^2. The high irradiance in a very small spot means that striking effects may be produced, even though the total power in the beam is modest.

A conventional lamp emitting total power P into 4π sr would produce irradiance I given by

$$I = \frac{4P}{\pi F^2} \tag{2.20}$$

when the light is focused by a lens with focal length F. This result is independent of lens diameter or the distance from the light source to the lens. Increasing the lens diameter would allow collection of more light but would increase the focal area by the same ratio because the effective beam divergence would be larger. The value of F cannot become very small, in practice not much less than 1 cm, because lenses of reasonable size with shorter focal length have very small F-number and will have large spherical aberration. Thus, Equation (2.20) predicts that to produce an irradiance equal to 10^7 W/cm^2 with a conventional light source, the source would have to emit total power around 10 MW, which is beyond the capability of any reasonable lamp. Therefore, a modest laser can produce irradiance greater than what may be obtained with very large conventional light sources.

The preceding description of the diffraction-limited focal spot size represents an ultimate limit that is set by the principles of optics. This limit may be approached with good engineering, that is, if the lens is free of aberrations and the spatial distribution in the beam is Gaussian. If the spatial distribution is more complicated, as is

often the case, or if the lens is less than perfect, one will not reach this ultimate limit. Typically, for a solid state laser with a complicated mode structure, the minimum focal spot size may be around 300 μm. To obtain a smaller size, one may aperture the beam so that the effective beam divergence is reduced. This involves a loss in total power and achieves no increase in radiance.

Despite their relatively poorer divergence angles, the beams from high-power solid state lasers can easily produce high irradiance. A focal area around 10^{-3} cm^2 is typical for a ruby laser beam focused with a simple lens. If the peak power is 10^6 W, then the irradiance will be 10^9 W/cm^2.

Gas lasers, which typically have better spatial profiles, may be focused more easily. Low-power gas lasers operating in the Gaussian mode may easily be focused to focal diameters in the 1–2 μm range, close to the diffraction limit. The depth of focus is proportionately small, of course. High-power gas lasers often operate in higher-order modes and cannot be focused to the diffraction limit. Typical values for focal diameter for high-power CO_2 lasers may be in the 100 μm range.

To make effective use of a laser, one often needs good focusing. One generally desires a small focal area in order to increase irradiance. To achieve a small focal area, one must use a lens with a short focal length, according to Equation (2.18). But this may be impractical in production applications because of limited depth of focus. One must provide sufficient depth of focus to compensate for vibration and for any inaccuracy in positioning of a workpiece. The depth of focus Z is approximately

$$Z \approx \frac{\lambda F^2}{a^2} \approx \frac{r_s^2}{\lambda} \tag{2.21}$$

where λ is the wavelength, a the radius of the lens, and F its focal length. Thus, the depth of focus increases with the square of the spot size. If the focal radius is around one wavelength, the depth of focus is also around one wavelength. One must use an optical system that provides a reasonable compromise between large depth of focus and small focal area.

Lens aberrations must also be considered; aberrations will degrade the performance of the focusing optics. For a monochromatic, collimated laser beam incident along the axis of the lens, many of the aberrations considered in elementary optics are unimportant. The most significant lens aberration is spherical aberration. Spherical aberration causes the light rays from a point source that enter the lens at different distances from its axis to be focused to a blurred circle, rather than a single point. Spherical aberration becomes more serious as the focal length of the lens decreases. This fact sets a practical lower limit to the focal length that can be used.

Spherical aberration can be minimized by use of best-shape lenses, by use of compound (multielement) lenses, or by use of aspheric lenses with specially ground shapes. In Chapter 5 we will discuss the choice of lenses in more detail, including compound and aspheric lenses. Here, we will describe only single spherical lenses. For use in the visible or near infrared portions of the spectrum, a planoconvex lens, with the convex side toward the laser, gives spherical aberration near the minimum.

In the far infrared spectrum, near 10 μm, the optimum lens shape is a meniscus (concave–convex) lens, again with the convex side toward the laser.

Figure 2-17 shows calculated results for focal spot diameter for a 1-in.-diameter lens with index of refraction equal to 1.5, used at a laser wavelength of 1 μm. The lower curve shows the diffraction-limited spot size in the absence of spherical aberration, as a function of the F-number of the lens. The upper curve shows the spot size determined by the limitation of spherical aberration, under the assumption that the lens has the optimum form to minimize spherical aberration. At small F-number, the effect of spherical aberration dominates, and the diffraction limit is not reached. In contrast to the prediction of Equation (2.19), the focal spot increases with decreasing F-number. The focusing may be improved by the use of air-spaced multielement lenses with short focal length. Under favorable conditions, one can indeed focus a laser beam to a spot of single-wavelength dimension. But in many practical industrial applications, one must retreat somewhat from this ultimate limit.

In Chapter 5, in the discussion on optics, we shall return to the topic of focusing and illustrate how the use of multielement lenses can reduce the effect of aberrations, so that the focal diameter is closer to the diffraction limit.

To summarize methods for focusing the beam to a small spot, first one should use a laser with a Gaussian beam profile. Then one should choose a lens with as short a focal length as possible, consistent with the desired depth of focus and the lens aberrations. The shape of the lens should be chosen to give minimum spherical aber-

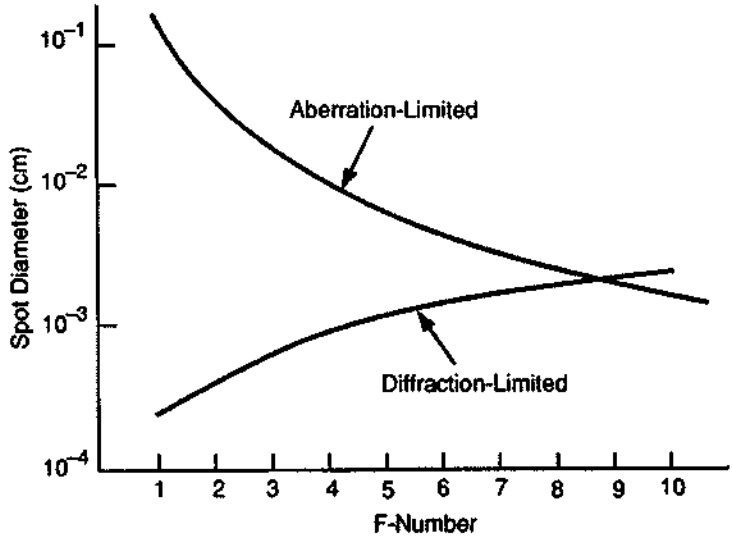


Figure 2-17 Calculated results for focal spot diameter as a function of the F-number of a focusing lens with 1 in. diameter and index of refraction 1.5, at a laser wavelength of 1 μm. The lower curve shows the diffraction-limited focal size in the absence of spherical aberration. The upper curve shows the focal size set by the limitations of spherical aberration for a best-form single-element lens.

ration, that is, planoconvex in the visible and near infrared and meniscus in the far infrared.

H. Power

The aspect of laser technology that has most caught the public imagination is the ability to produce very high levels of output power. The highest levels of power come from pulsed lasers (such as Nd:glass) and are achievable only in short pulses.

We should distinguish between experimental devices and commercially available lasers. Experimental lasers usually have produced higher levels of power than the same type of laser that can be purchased from a commercial manufacturer. We shall describe the output power characteristic of different laser types in Chapter 3. Here, we shall merely note some rough numbers. The numbers given here are upper limits; most practical lasers operate at lower power levels. For continuous output, the highest values are produced by CO_2 lasers. Large experimental CO_2 lasers have exceeded hundreds of kilowatts, and commercial continuous CO_2 lasers around 25 kW are available. Among pulsed lasers, solid state devices have produced the highest powers. Commercial Q-switched Nd:glass lasers with peak power up to 10^{11} W are available (with nanosecond pulse duration). Experimental Nd:glass devices for laser-assisted thermonuclear fusion have reached levels around 10^{14} W.

References

[1] A. L. Schawlow and C. H. Townes, *Phys. Rev.* **112**, 1940 (1958).
[2] T. S. Jaseja, A. Javan, and C. H. Townes, *Phys. Rev. Lett.* **10**, 165 (1963).
[3] J. Davit and C. Charles, *Appl. Phys. Lett.* **22**, 248 (1973).
[4] J. E. Bjorkholm and R. H. Stolen, *J. Appl. Phys.* **39**, 4043 (1968).
[5] J. F. Ready, *Effects of High Power Laser Radiation*. Academic Press, New York, 1971.
[6] R. J. Collier, C. B. Burckhardt, and L. H. Lin, *Optical Holography*. Academic Press, New York, 1971, Equation (7.17).
[7] D. Chen and J. F. Ready, *Bull. Am. Phys. Soc.* **11**, 454 (1966).
[8] D. F. Nelson and R. J. Collins, *J. Appl. Phys.* **32**, 739 (1961).
[9] G. Magyar, *Opto-Electron.* **2**, 68 (1970).
[10] D. A. Berkley and G. J. Wolga, *Phys. Rev. Lett.* **9**, 479 (1962).

Selected Additional References

A. Linewidth

A. L. Bloom, *Gas Lasers*, Wiley, New York, 1968, Chapters 3 and 4.
G. D. Boyd and H. Kogelnik, Generalized Confocal Resonator Theory, *Bell. Syst. Tech. J.* **41**, 1347 (1962).
R. C. Duncan, Z. J. Kiss and J. P. Wittke, Direct Observation of Longitudinal Modes in the Output of Optical Masers, *J. Appl. Phys.* **33**, 3568 (1962).
K. D. Mielenz *et al.*, Reproducibility of Helium–Neon Laser Wavelengths at 633 nm, *Appl. Opt.* **7**, 289 (1968).

B. Collimation

R. J. Collins *et al.*, Coherence, Narrowing, Directionality and Relaxation Oscillations in the Light Emission from Ruby, *Phys. Rev. Lett.* **5**, 303 (1960).
D. R. Herriott, Optical Properties of a Continuous Helium–Neon Optical Maser, *J. Opt. Soc. Am.* **52**, 31 (1962).
D. C. Sinclair and W. E. Bell, *Gas Laser Technology*, Holt, New York, 1969.

C. Spatial Profiles of Laser Beams

A. L. Bloom, *Gas Lasers*, Wiley, New York, 1968, Chapter 3.
G. D. Boyd and H. Kogelnik, Generalized Confocal Resonator Theory, *Bell. Syst. Tech. J.* **41**, 1347 (1962).
E. S. Dayhoff and B. Kessler, High-Speed Sequence Photography of a Ruby Laser, *Appl. Opt.* **1**, 339 (1962).
V. Evtuhov and J. K. Neeland, Observations Relating to the Transverse and Longitudinal Modes of a Ruby Laser, *Appl. Opt.* **1**, 517 (1962).
H. Kogelnik and T. Li, Laser Beams and Resonators, *Appl. Opt.* **5**, 1550 (1966).
H. Kogelnik and W. W. Rigrod, Visual Display of Isolated Optical Resonator Modes, *Proc. IRE* **50**, 220 (1962).

D. Temporal Behavior of Laser Output

R. J. Collins *et al.*, Coherence, Narrowing, Directionality and Relaxation Oscillations in the Light Emission from Ruby, *Phys. Rev. Lett.* **5**, 303 (1960).
A. J. DeMaria, D. A. Stetser, and H. Heynau, Self Mode-Locking of Lasers with Saturable Absorbers, *Appl. Phys. Lett.*, **8**, 174 (1966).
A. J. DeMaria *et al.*, Picosecond Laser Pulses, *Proc. IEEE* **57**, 2 (1969).
M. A. Duguay, S. L. Shapiro, and P. M. Rentzepis, Spontaneous Appearence of Picosecond Pulses in Ruby and Nd:Glass Lasers, *Phys. Rev. Lett.* **19**, 1014 (1967).
G. Kachen, L. Steinmetz, and J. Kysilka, Selection and Amplification of a Single Mode-Locked Optical Pulse, *Appl. Phys. Lett.* **13**, 229 (1968).
F. J. McClung and R. W. Hellwarth, Characteristics of Giant Optical Pulsations from Ruby, *Proc. IEEE* **51**, 46 (1963).

E. Coherence

N. P. Barnes and I. A. Crabbe, Coherence Length of a Q-Switched Nd:YAG Laser, *J. Appl. Phys.* **46**, 4093 (1975).
M. Born and E Wolf, *Principles of Optics*, 6th ed., Pergamon, Oxford, 1980.
M. S. Lipsett and L. Mandel, Coherence Time Measurements of Light from Ruby Optical Masers, *Nature (London)* **199**, 553 (1963).
D. C. W. Morley *et al.*, Spatial Coherence and Mode Structure in the HeNe Laser, *Brit. J. Appl. Phys.* **18**, 1419 (1967).

F. Radiance

W. F. Hagen, Diffraction-Limited High-Radiance Nd:Glass Laser System, *J. Appl. Phys.* **40**, 511 (1969).
D. C. Hanna, Increasing Laser Brightness by Transverse Mode Selection—1, *Opt. Laser Technol.*, p. 122 (August 1970).
G. Magyar, Simple Giant Pulse Ruby Laser of High Spectral Brightness, *Rev. Sci. Instrum.* **38**, 517 (1967).

G. Focusing Properties of Laser Radiation

A. L. Bloom, *Gas Lasers,* Wiley, New York, 1968, Chapter 4.
W. H. Lowrey and W. H. Swantner, Pick a Laser Lens that Does What You Want It to, *Laser Focus World,* p. 121 (May 1989).
R. T. Pitlak, Laser Spot Size for Single-Element Lens, *Electro-Opt. Syst. Design,* p. 30 (September 1975).

H. Power

Laser Focus World Buyers Guide, PennWell Publishing Co., Nashua, NH (1995).

Chapter 3 | Practical Lasers

In this chapter we shall describe the status of well-developed laser types and discuss the lasers that have reached commercial status. By commercial lasers, we mean those available for sale by a number of different suppliers. The properties that we shall describe in this chapter do not necessarily represent optimum values that have been obtained in laboratory experiments; rather, we shall emphasize what one can buy readily from commercial suppliers.

We shall discuss a variety of different laser materials. This is a relatively small set of materials taken from the large class of hundreds of materials in which laser action has been demonstrated in the laboratory. We describe the set of materials that work well as lasers, or at least the ones that are most fully developed.

Future development will undoubtedly change the list of important types of lasers. New and emerging developments offer promise for new types of lasers operating at different wavelengths. In the future, these may well supplant some of the lasers that are now commonly used. In the next chapter we shall discuss some of the prospects for new laser types.

A. Gas Lasers

Lasers that utilize a gaseous material as the active laser medium have become extremely common. There are compelling reasons for the use of a gas. The volume of active material can be large, in contrast to the cases of crystals and semiconductors. The material is relatively inexpensive. There is no possibility of damage to the active medium, in contrast to the case of high-power solid lasers. The medium is homogeneous, so that one avoids problems of inhomogeneity common in solid lasers. Heat can be removed readily by transporting the heated gas out of the region of laser operation. In some types of lasers, high power outputs are aided by flowing of the gas.

We shall discuss several different types of gas lasers. These include the neutral gas laser (typified by the helium–neon laser), the ionized gas laser (typified by ar-

gon), the molecular gas laser (typified by carbon dioxide), the excimer laser (typified by krypton fluoride), and the metal vapor laser (typified by copper vapor).

1. NEUTRAL GAS LASERS

Laser operation involving neutral gas species occurs in the positive column of a glow discharge in a medium such as a mixture of helium and neon. The excitation mechanism for a helium–neon laser has been discussed in Chapter 1. Briefly, the passage of an electrical current, with a current density around 100 mA/cm^2, leads to excitation of the helium atoms through collisions with electrons in the electrical discharge. This excitation energy in turn is transferred through atomic collisions to the upper laser levels of the neon atoms, thus producing the population inversion.

The first neutral gas laser was demonstrated in 1961, with the operation of the 1.15 μm line of the helium–neon laser. This was followed in 1962 with the operation of the familiar 633 nm orange-red line, which remains one of the most common laser emissions. Laser operation has been demonstrated in many other neutral gases and gas mixtures. The most notable ones involve the noble gases, neon, argon, krypton, and xenon, with a variety of wavelengths available through the visible and near infrared portions of the spectrum. No other neutral gas laser is nearly as common as the helium–neon laser.

The orange-red helium–neon laser is perhaps the most familiar of all lasers. It has been estimated that more than 400,000 helium–neon lasers are manufactured and sold each year. They represent a compact, portable, and easily usable source of continuous, visible laser radiation and have many uses for demonstration and educational purposes.

The energy level diagram for the helium–neon laser has been shown in Chapter 1. The helium is excited by the passage of an electrical current through the gas mixture. The upper laser levels of the neon atoms are populated through collisions with excited helium atoms. Because of the desirability of visible operation, most helium–neon lasers are constructed to operate at the 632.8 nm wavelength. We see from the energy level diagram (Figure 1-10) for neon that there is competition between the 633 nm line and other lines. Thus, if the 1.15 μm line is operating, it tends to fill the terminal level for the 632.8 nm line. Similarly, if the 3.39 μm line is operating, it uses up the excitation that would otherwise contribute to the 632.8 nm emission. Thus, one usually tries to suppress operation of the infrared lines. The 1.15 μm line may be suppressed by the use of mirrors that have low-reflectivity at that wavelength. However, the 3.39 μm line has very high gain and is difficult to suppress with low-reflectivity mirrors alone. Many helium–neon lasers incorporate magnets that provide an inhomogeneous magnetic field that varies widely over different parts of the gas mixture. This inhomogeneous magnetic field broadens the profile of the 3.39 μm transition and reduces the gain available for it, thus favoring the operation at 633 nm.

In addition, operation at other visible laser wavelengths may be obtained by insertion of a prism in the helium–neon laser. The operation proceeds from the same

upper level as the 632.8 nm line, but to different lower sublevels. The dispersion of the prism causes the different wavelengths to travel in slightly different directions. The prism is mounted so that only one of the wavelengths will follow a reentrant path through the laser cavity. Visible wavelengths that may be selected in this way include 543, 594, 604, and 612 nm. Infrared helium–neon lasers operating at 1.15, 1.52, and 3.39 μm are also available.

The gain of the helium–neon laser at 632.8 nm is relatively low, so that only very small losses can be tolerated in the laser cavity. Thus, the mirrors must be of high quality with low losses. The output mirror typically can have transmission of no more than 1 or 2 percent. If the mirrors are external to the gas–filled tube, the tube must have Brewster angle windows.

In the helium–neon laser, the gain is an inverse function of the diameter of the gas tube, because the lower level of the laser transition is depopulated by collisions of the neon atoms with the walls of the tube. In order that the collision rate of neon atoms with the wall be high, the tube diameter must be small. A typical tube diameter is around 1 mm. This fact has practical consequences. It limits the power output from helium–neon lasers because the volume of gas cannot be readily increased. In addition, it affects the beam divergence angle, because the walls of the tube define a limiting aperture. According to Equation (2.6), the typical beam divergence angle for a helium–neon laser will be a few tenths of a milliradian.

The gas pressure in a helium–neon laser is usually in the region of a few Torr, with the mixture consisting typically of 90 percent helium and 10 percent neon. Early work involved many studies of the output of helium–neon lasers as a function of gas pressure, gas mixture, tube diameter, tube current, and so forth. Commercial helium–neon lasers are constructed to provide optimum output characteristics.

The helium–neon laser represents very well developed mature technology, with many models available from a number of commercial manufacturers. The lifetime of helium–neon lasers exceeds 20,000 hours, according to manufacturers' estimates. Table 3–1 summarizes some of the typical parameters available commercially.

The range of output power extends from a few tenths of a milliwatt to 35 mW in devices that are not frequency stabilized. Such lasers have amplitude stability around 0.5 percent over short periods of time and power drift of several percent over longer times (hours).

Most commercial helium–neon lasers are constructed to operate in the TEM_{00} mode. This is accomplished by ensuring that the ratio of the length of the tube to its diameter favors the TEM_{00} mode. Other, higher-order, modes have larger diameters and cannot operate within the narrow aperture defined by the tube.

Figure 3–1 shows the detailed structure of a typical small commercial helium–neon laser. The design is the result of extensive engineering development. It is compatible with mechanized production and rapid packaging. The seals that attach the mirrors to the tube usually are formed by ceramic frits, which are baked at high temperature to ensure cleanliness and to avoid contamination of the gas. Such seals are referred to as "hard seals" and are accepted current practice in laser fabrication. Be-

Table 3-1 **Typical Parameters of Commercial Helium–Neon Lasers**

Lasers operating at 632.8 nm:
Power: 0.5–35 mW
Mode: TEM_{00} (95%)
Beam diameter: 0.5–1.5 mm at $1/e^2$ points (typical)
Beam divergence: 1–2 mrad (diffraction-limited)
Power drift: 5% (over 8 hours)
Amplitude noise: 0.5% (to 100 KHz)
Lifetime: 20,000 hours
Options
Random polarization or linear polarization (polarization ratio of selected component typically 500 to 1)
Alternate wavelengths: 0.543, 0.594, 0.612, 1.15, 1.52, 3.39 μm
Single longitudinal mode: ±3 MHz over 8 hours, 1 mW

cause the lifetime of the helium–neon laser is often limited by leakage of helium from the tube, the gas reservoir extends the lifetime of the laser.

A helium–neon laser without frequency stabilization may have several longitudinal modes present in the output. Figure 3-2 shows a drawing of the frequency spectrum of a typical helium–neon laser, as obtained with a spectrum analyzer. In this case, four longitudinal modes are present. For a 50-cm-long tube, the spacing between the modes, $c/2D$, will be 300 MHz. The total linewidth will then be about 900 MHz. We note that the total fluorescent linewidth of the 632.8 nm helium–neon laser transition is approximately 1700 MHz, so that the emission covers about half

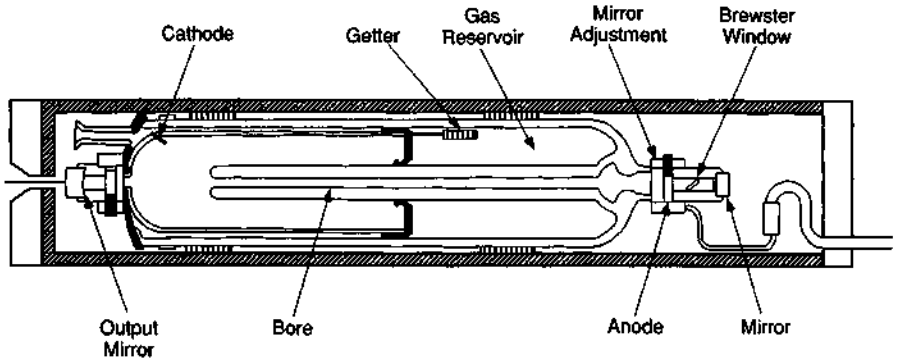


Figure 3-1 Structure of a helium–neon laser designed for volume production. (Figure courtesy of Melles Griot.)

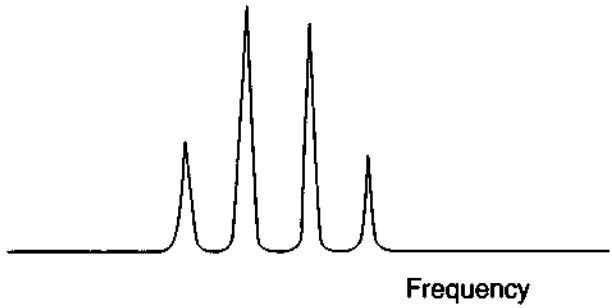


Figure 3-2 Frequency spectrum of a typical unstabilized helium–neon laser.

of the fluorescent linewidth. The coherence length of a laser with four modes present will be only 33 cm, according the concepts developed in Chapter 2.

Commercial frequency-stabilized helium–neon lasers are available and are used in applications like interferometry, which require high coherence. These lasers utilize a temperature-stabilized oven and feedback control to keep the laser operating at the center of the Lamb dip, according to the discussion in Chapter 2. The frequency stability of such lasers is of the order of a few megahertz, corresponding to a coherence length around 100 m.

In summary, the well-developed helium–neon laser, with a reddish-orange collimated output in the milliwatt range, has become a familiar device in common use for many applications, including interferometry, holography, velocimetry, alignment, distance measurement, and barcode reading.

2. *IONIZED GAS LASERS*

An ion laser consists of a plasma with a glow discharge at high current density passing through it. The most common ionized gas laser utilizes ionized argon as the laser medium. A partial energy level diagram relevant to the argon ion laser is shown in Figure 3-3. A number of laser transitions are indicated. The strongest laser transitions are at 488 and 514.5 nm.

Unlike the helium–neon laser, for which the energy levels are those of neutral atoms, the energy levels in the figure are those of singly ionized argon gas. The lowest energy level shown in the figure is the ground state of the argon ion, which lies 16 eV above the ground state of the neutral argon atom. In addition, the upper energy levels lie about 20 eV above the ionic ground state. Thus, a considerable amount of energy must be supplied to the neutral argon atom to raise it to the upper laser level for laser operation.

The argon laser has higher gain than the helium–neon laser, and much larger amounts of power may be extracted from it. The output power scales nonlinearly with the current density, so that it is desirable to operate argon lasers with a narrow bore and a high electrical current. Current densities above 100 A/cm^2 may be em-

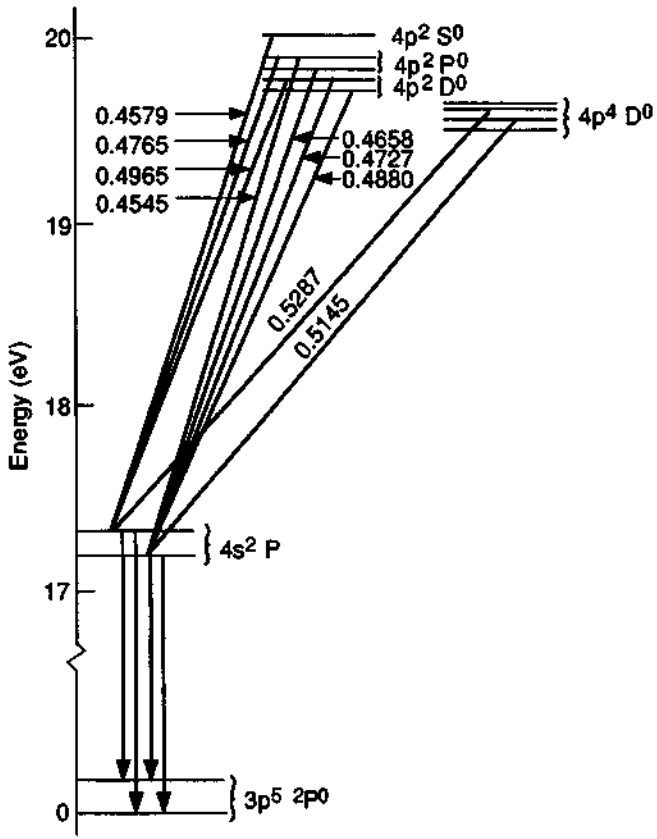


Figure 3-3 Partial energy level diagram of singly ionized argon, showing energy levels relevant to the operation of the argon ion laser. The spectroscopic notation for the energy levels is indicated. Various laser wavelengths, in micrometers, are indicated.

ployed in argon lasers. The high current density produces heating and erosion of the plasma tube walls, and thus it strongly influences the design of argon lasers.

Materials used to define the bore of an argon laser must be resistant to both high temperature and to sputtering by the electrical discharge. A common design involves use of electrically isolated, radiation-cooled annular segments of material to confine the discharge, enclosed in a vacuum envelope. The material that is most commonly used now for the segments is tungsten. The envelope is usually alumina ceramic or quartz. Argon lasers require active cooling. Air cooling may be used in relatively low power models, but water cooling is employed at higher levels of power.

A typical design for a segmented bore argon laser tube is shown in Figure 3-4. Each bore segment has a hole in its center to define the discharge region, and holes around the central region to allow for gas return. The figure also shows the wavelength

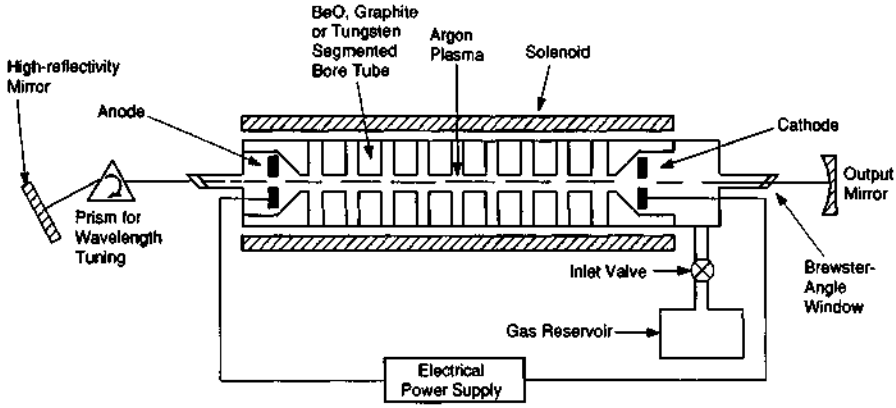


Figure 3-4 Typical argon laser structure with segmented bore.

selecting prism, which allows operation on a single one of the wavelengths available. The prism is rotated to select the desired wavelength. Without the prism, multiple wavelengths are present in the output simultaneously, and the laser is said to operate multiline. The solenoid supplies a longitudinal magnetic field, which confines the electrical current and increases the current density.

The argon gas is depleted in the course of operation, so that gas refill reservoirs are included in the design of many argon lasers. When the argon pressure drops below a specified value, the operator activates the refill system to restore the pressure to the specified value.

Characteristics of commercially available argon lasers are presented in Table 3-2. The output power available ranges from a few tenths of a watt to 25 W. The beam is usually a diffraction-limited TEM_{00} mode. The output power of an argon laser is usually specified in terms of a multiline output, that is, the sum of the powers in all the wavelengths that are simultaneously present in the output (488, 514.5 nm, etc.). Table 3-3 indicates the split between the different wavelengths present in a nominal 20-W argon laser.

Single-line operation is possible if one adds a wavelength selecting element, such as a prism, as was shown in Figure 3-4, to the laser cavity. Because of the dispersion of the prism, only one wavelength will be incident normally on the mirrors, and so only that wavelength will operate. Selection of any particular wavelength is accomplished by rotating the prism. The output power will be reduced from the multiline power, to a value comparable with the contribution of the selected line to the multiline output. Thus, if a multiline 20-W argon laser is constrained to operate at 514.5 nm, the output power may be only slightly greater than the 8.5 W indicated in Table 3-3.

Argon lasers offer outputs in the 334 to 364 nm wavelength range. This makes the argon laser one of the few commercial sources of continuous ultraviolet laser light.

Table 3-2 **Commercial Ion Laser Characteristics**

Argon
Multiline operation: 514.5, 488, 476.5, 501.7 nm, etc.
Ultraviolet operation in 334 to 364 nm range
Single line operation with wavelength selecting element
Power to 25 W continuous TEM_{00} multiline, to 10 W at 514.5 nm, to 7 W in multiline ultraviolet option
Usually TEM_{00} mode
Beam diameter: 0.5–1.5 mm typical
Beam divergence: 0.5–2 mrad typical (diffraction-limited)
Power stability: ±0.2% typical with feedback stabilization
Relatively high electrical power and cooling water requirements at high power
Krypton
Multiline operation: 520–676 nm
Power to 16 W continuous TEM_{00} multiline, to 3.5 W at 674.1 nm
Mode: TEM_{00} usually
Other parameters similar to argon
Helium–cadmium
Output at 442 and 325 nm
Power to 215 mW continuous multimode at 442 nm, to 170 mW TEM_{00}
Mode: TEM_{00} usually
Applications
Test and measurement
Entertainment
Research and development
Material processing
Lithography

The requirement for large electrical current and the rather complicated structure of the tube make the argon laser more prone to failure than the long-lived helium–neon laser. The manufacturers' warranties for argon lasers are shorter than for helium–neon lasers. The operating lifetimes of argon lasers before failure are typically several thousand hours.

Krypton ion lasers are very similar in construction and characteristics to argon lasers. Table 3-2 includes some of the characteristics of krypton lasers. Table 3-3 presents the split between the power in the different lines of a nominal 14-W krypton laser in multiline operation. In single-line operation, the krypton laser provides somewhat lower output power than an argon laser. The main reason for choosing a krypton laser instead of an argon laser would be to obtain a different variety of

Table 3-3 **Representative Ion Laser Output**

Nominal 20-W multiline argon laser		Nominal 14-W multiline visible krypton laser	
Wavelength (nm)	**Power (W)**	**Wavelength (nm)**	**Power (W)**
582.7	1.4	676.4	0.9
514.7	8.5	674.1	3.5
501.7	1.4	568.2	1.1
496.5	2.4	530.9	1.5
488.0	6.5	520.8	0.7
476.5	2.4	482.5	0.4
472.7	1.0	476.2	0.4
465.8	0.6	468.0	0.5
457.9	1.4	415.4	0.3
454.5	0.6	413.1	1.8
333.6–363.8[a]	4.0	406.7	0.9
		337.5–356.4[a]	2.0

[a]With special mirrors.

wavelengths, particularly wavelengths in the red and yellow portions of the spectrum, at power levels above those of the helium–neon laser. Mixed gas lasers containing both argon and krypton are available and provide a composite choice of wavelengths. Such mixed gas lasers provide lines in the red, green, and blue portions of the spectrum and thus are suitable for full color displays.

The helium–cadmium laser is also an ion laser, because its laser operation utilizes the energy levels of ionized states of gaseous cadmium. In many of its characteristics, however, it is similar to the metal vapor lasers to be described later. The cadmium is vaporized by a heater. The excitation of the upper laser levels of the vaporized cadmium is produced similarly to the excitation of the neon levels in the helium–neon laser. In the electric discharge through the helium–cadmium gas mixture, helium atoms are excited by collisions with electrons. The excited helium atoms then collide with ground state cadmium atoms. The collisions produce excited levels of the cadmium ion. The process is favorable for producing populations for two different laser lines, at wavelengths of 441.6 and 325 nm.

Development of the helium–cadmium laser represented a great technical challenge because of the presence of a hot metallic vapor. Successful operation of the helium–cadmium laser requires production of a sufficiently high vapor pressure of cadmium metal, good control of the vapor in the discharge, prevention of chemical reactions between the hot cadmium vapor and other tube components, and avoidance of deposition of cadmium metal on the optical surfaces. A substantial amount of engineering work has led to simple, reliable, and compact commercial helium–cadmium lasers. Operating lifetimes greater than 6000 hours have been quoted.

Some characteristics of helium–cadmium lasers are presented in Table 3-2. In commercial devices, one may obtain output power over 200 mW multimode and up to 170 mW in the TEM_{00} mode at 441.6 nm and up to 50 mW mutimode at 325 nm.

Small helium–cadmium lasers emitting a few milliwatts of power at 441.6 nm require about 160 W of input power, whereas larger models, operating in the 100 mW range, require several hundred watts of input power. Heating of the cadmium metal to provide the necessary vapor pressure may require a few tens of watts of additional electrical power. Most commercial helium–cadmium lasers require 120 VAC wall plug power, but some higher-power models may use 220 VAC.

The helium–cadmium laser represents a compact continuous source of visible laser light at relatively low levels of power and can be used for applications similar to those of the helium–neon laser, except at a shorter wavelength. The 325 nm line has been useful for lithographic applications.

One final ion laser that has some commercial availability is the xenon ion laser, which operates in the blue and green portions of the spectrum. Usually operated pulsed at high current density, this laser emits several lines in the blue and green regions from 480 to 540 nm. Although the pulse energy tends to be low, the peak power may be high enough to allow evaporation of thin films and some microelectronic processing applications.

3. MOLECULAR LASERS

So far, all the lasers discussed operate using energy levels that are characteristic of electronic transitions. These energy levels correspond to transitions of an electron from one of its energy levels in an atom or ion to another energy level available to it. In addition, there are other important energy levels that involve no change in electronic state. These are the vibrational and rotational energy levels of molecules. A polyatomic molecule may have a number of internal degrees of freedom that involve vibration and rotation of the molecule. According to the laws of quantum mechanics, the energies associated with these vibrations and rotations are quantized, so that there is a spectrum of discrete energy levels corresponding to them. If a population inversion is produced between vibrational or rotational levels, laser action may occur. Typically, the energies associated with electronic transitions are larger than the energies associated with vibrational transitions, which in turn are larger that the energies associated with rotational transitions. Thus, molecular lasers that utilize these energy levels tend to operate in the far infrared portion of the spectrum. The vibrational and rotational energy levels that are utilized are sublevels of the ground electronic state of the molecule.

Carbon Dioxide Lasers

The most important molecular laser uses CO_2 gas as its active medium. The CO_2 molecule is a linear molecule, with the three atoms arranged in a straight line with the carbon atom in the middle. There are three different types of vibration that can

occur in this molecule. These vibrations are illustrated schematically in the top portion of Figure 3-5. The arrows indicate the instantaneous displacements of the atoms from their equilibrium positions, which are denoted by the circles. The motion is a simple harmonic motion, with all the atoms moving in phase and oscillating about their equilibrium positions. In the first mode of vibration, the carbon atom remains stationary and the oxygen atoms move in opposite directions along the line of symmetry, as indicated by the arrows. In the second mode of vibration, which is a bending motion, all the atoms move in a plane perpendicular to the line of symmetry. The carbon atom moves in one direction, while the oxygen atoms move in the opposite direction. The third mode is an asymmetric mode in which all the atoms move on the same line. At any time, the carbon atom is moving in a direction opposite to the oxygen atoms.

Energy is associated with these vibrational modes, and in accordance with quantum mechanics, the energy is quantized. In the ground vibrational state, there are no quanta of vibration. Excited states correspond to the presence of one or more quanta of vibrational energy. The molecule can vibrate in more than one mode at a time and can have more than one quantum in each mode. The notation adopted to denote the energy levels is of the form (ij^lk), where i, j, and k are integers denoting the quanta of vibration in each of the three modes. The integer i corresponds to the symmetric vibration, j to the bending vibration, and k to the asymmetric vibration. The superscript l is an additional quantum number, which arises because the bending vibration can occur in either of two perpendicular senses. Thus, the level (10^00) represents a state with one quantum of symmetric vibration and no quanta in either of the other two modes.

In addition, each of the vibrational levels is split into a number of rotational sublevels. The rotational energy is small compared with the vibrational energies. A simplified energy level diagram showing the vibrational energy levels of greatest importance for the operation of the CO_2 laser appears in the bottom portion of Figure 3-5. We note explicitly that these are all sublevels of the ground electronic state of the CO_2 molecule, and that rotation of the molecule further splits the vibrational levels into sublevels.

The upper laser levels of the CO_2 molecule are excited by collisions with excited nitrogen molecules, which in turn are excited by collisions with electrons in an electrical discharge through a carbon dioxide–nitrogen mixture. The (00^01) level of the CO_2 molecule is only 18 cm^{-1} above the first excited vibrational level of the nitrogen molecule. Energy is transferred in collisions from nitrogen molecules to CO_2 molecules, raising them to the (00^01) level. This produces a population inversion between the (00^01) level and the (10^00) and (02^00) levels. These correspond to laser operation at 10.6 and 9.6 μm, respectively. The gain is higher for the 10.6 μm transition, so that is the preferred transition. Because the two laser transitions share a common upper level, operation at 10.6 μm usually precludes operation at 9.6 μm. The laser will usually operate at 10.6 μm unless a wavelength selecting mechanism, like a grating, is employed. The lower laser levels relax by collisions and radiative

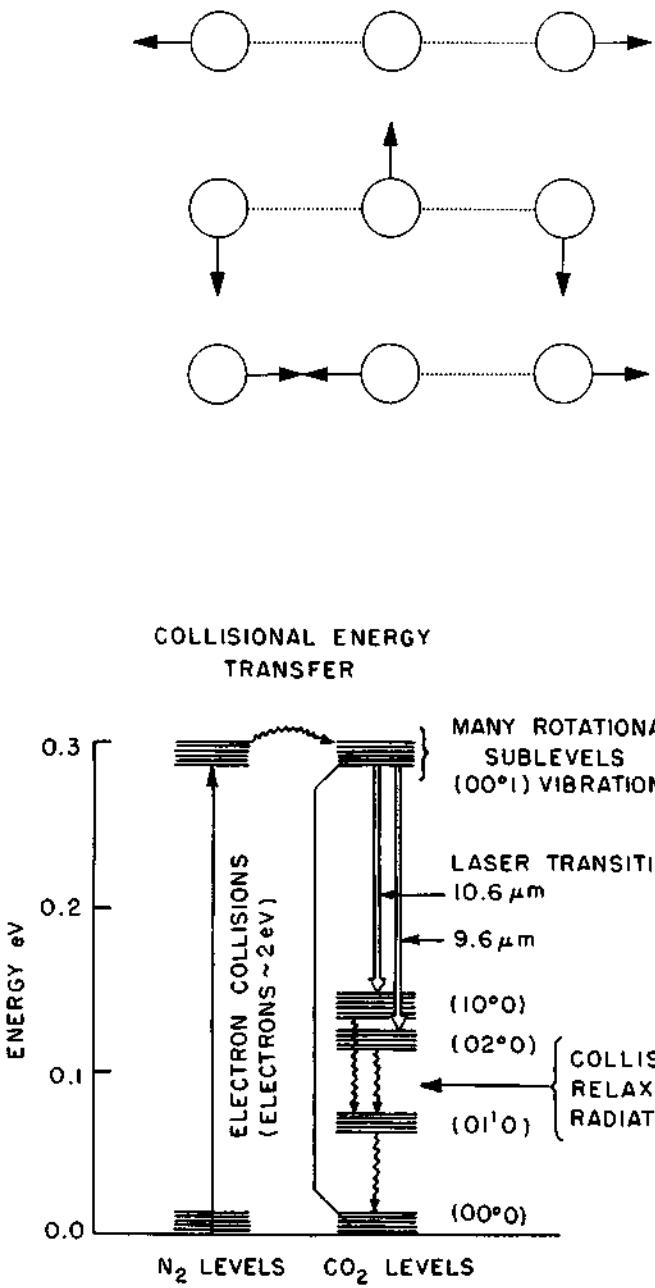


Figure 3-5 Top: Schematic diagram of the vibrations of the carbon dioxide molecule. Bottom: Energy levels relevant to the operation of a CO_2–N_2 laser. The numbers for the vibrational states are defined in the text.

emission, eventually returning the molecule to the ground ($00^0 0$) level. The coupling of the vibrational and rotational motions of the CO_2 molecule leads to considerable additional substructure. Transitions may occur between any levels for which the change ΔJ in rotational quantum number J is $\Delta J = \pm 1$. Thus, the output of the laser may have many possible different lines. These lines are denoted P or R, corresponding to ΔJ equal to +1 or −1, respectively, followed by the rotational quantum number of the lower rotational level. Thus, the notation $P(22)$ refers to a line for which the upper level has $J = 21$ and the lower level has $J = 22$. Figure 3-6 shows some details of the rotational structure.

Figure 3-7 shows the gain distribution for the possible lines in the 10–11 μm region, as a function of the rotational quantum number of the lower level. The separation of the P and R branches and the multiline nature of the emission are clear. Usually, the output of a CO_2 laser will consist of one or two of these lines, most often $P(20)$ or $P(22)$. Tuning to any of the lines in the figure is possible if one uses a wavelength selecting element, like a rotatable diffraction grating in place of the high-reflectivity mirror, although the power is reduced compared with operation on the $P(20)$ or $P(22)$ lines.

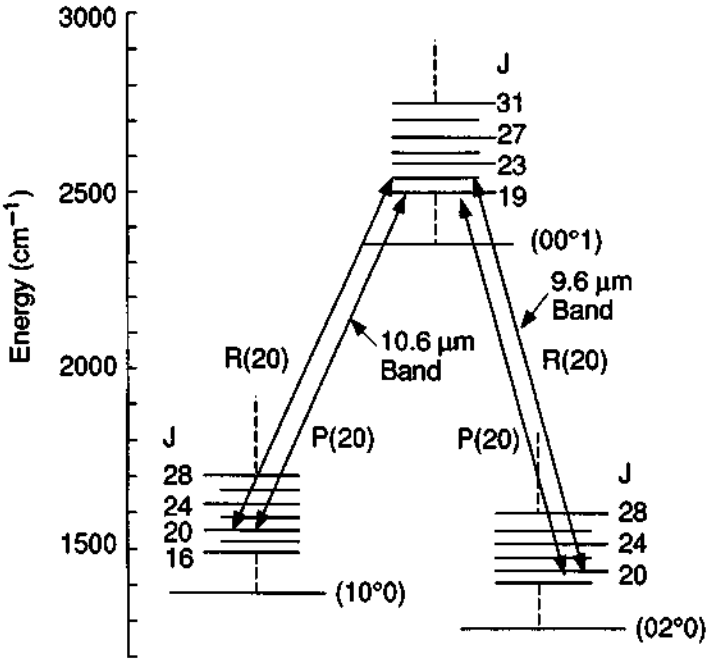


Figure 3-6 Rotational substructure of some CO_2 vibrational energy levels. The notation is defined in the text.

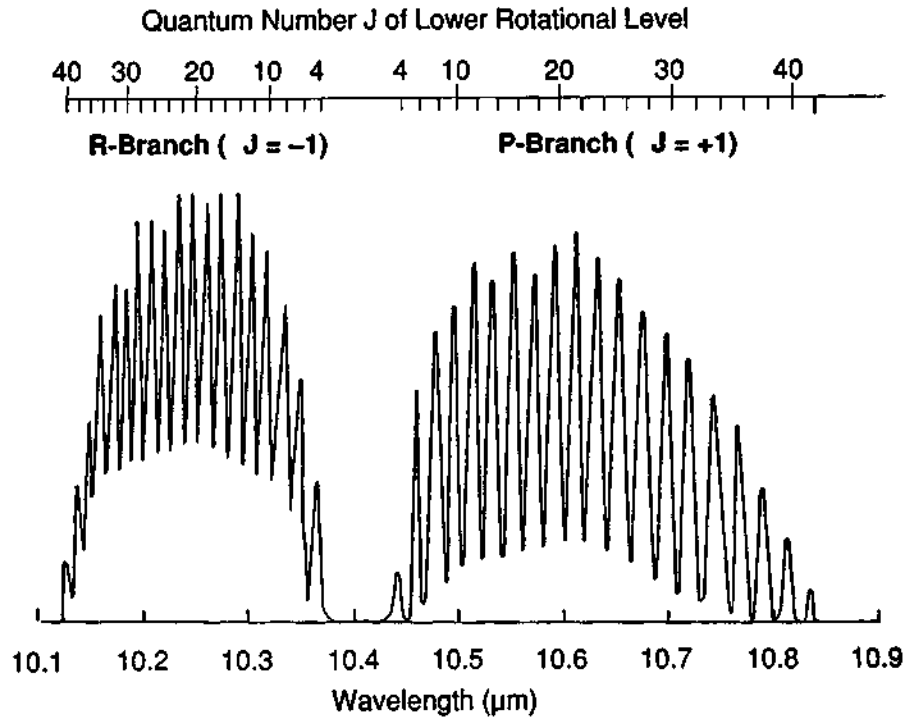


Figure 3-7 Gain distribution in CO_2 as a function of wavelength and the quantum number J of the lower vibrational level. (From U. P. Oppenheim and A. P. Devir, *J. Opt. Soc. Am.* **58**, 585 (1968).)

As the lower portion of Figure 3-5 showed, the population inversion in the CO_2 laser is produced by collisions with nitrogen molecules. Early experiments showed laser operation in pure CO_2, but the power was low. High power was obtained by increasing the efficiency of excitation of the upper laser level through the addition of nitrogen. The addition of helium to the gas mixture further increases the output power. Helium acts to deplete the population of the lower laser level by collisions and helps transfer heat to the walls because of the high mobility of helium atoms. Practical CO_2 lasers thus use a mixture of carbon dioxide, nitrogen, and helium, although they are generally referred to simply as CO_2 lasers.

We note that CO_2 lasers generally are flowing gas systems. The gas flows through the system and is exhausted. In a sealed CO_2 laser, the interaction of the electrical discharge with the gases causes decomposition, buildup of harmful products, and degradation of the laser output over a period of time. This is avoided simply by continuously exhausting the gas. For most industrial applications, the weight of the gas supply and the cost of the gas are not dominant factors, so that flowing systems are

used. Fast flow of the gas provides convective cooling and leads to the possibility of high output powers from relatively small volumes of gas.

In situations where flowing the gas is not feasible, sealed-off CO_2 lasers have been designed to last thousands of hours. Such lasers involve the use of getter materials that will remove the harmful decomposition products.

The CO_2 laser is also notable for relatively high efficiency of conversion of electrical energy into optical energy, in contrast to most lasers, which have low efficiency. Typical laser efficiencies are in the range of a few tenths of one percent to a few percent. The CO_2 laser is capable of operation in the range of 10–30 percent. We note that the operating conditions that yield the highest efficiency do not necessarily yield the highest power.

Carbon dioxide lasers most often operate continuously, but they may readily be pulsed by pulsing the electrical power supply. In that case, the pulse is of the form shown in Figure 2-13, with an initial spike followed by a longer lower-power tail, at a power level close to what is available in continuous operation. This pulse form is often useful for metalworking, with the initial spike breaking down the metal surface so that the energy in the tail can couple effectively into the metal.

Early studies of CO_2 laser operation indicated that the power output scaled with the length of the laser tube. Typical power extraction was around 80 W/m of discharge length, for CO_2 lasers with relatively slow gas flow. Thus, continuous output power of a few hundred watts was easy to obtain in devices of reasonable size. But an output in the kilowatt range required an unreasonably long tube. In the late 1960s, it was discovered that increasing the rate of gas flow increased the output power. The results of one study are presented in Figure 3-8. The power increased rapidly with gas flow rate until it saturated at a flow near 4000 ft^3/min. With fast gas flow, it is possible to obtain output powers of several hundred watts per meter of discharge length. It then becomes possible to construct lasers with high power (kilowatts) in devices of reasonable size.

Thus, the evolution of CO_2 laser construction has been as illustrated in Figure 3-9. The early models of CO_2 lasers were constructed as shown in the top portion of the figure, with gas flow along the axis of the tube sufficient to remove decomposition products. Such lasers are now referred to as slow axial flow devices. A second class of lasers, with capability of higher power, utilizes fast flow of the gas along the length of the tube. Such devices are called fast axial flow lasers. Such devices can emit perhaps 600 W/m of discharge length. Then models were developed in which the gas flow is perpendicular to the long axis of the laser, that is, transverse to the optical path, as shown in the bottom portion of Figure 3-9. Because the impedance to gas flow is reduced with transverse flow, the power output can be increased further. Also, the electrical discharge may be perpendicular to the optical path. It becomes easier to supply the optimum electric field (voltage per unit length) when the length of the discharge is reduced. Thus, in many modern lasers, the electric field is also supplied perpendicular to the optical path. With a configuration such as that shown an the bottom portion of the Figure 3-9, extraction of several kilowatts of continuous power in a reasonably sized device becomes possible. Lasers constructed

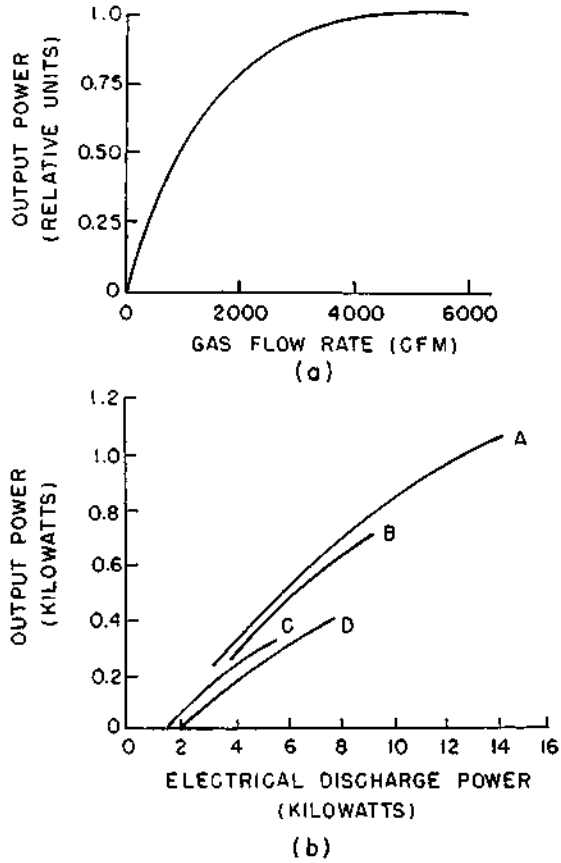


Figure 3-8 Output of fast gas flow CO_2 laser. Part (a) shows the output as a function of gas flow rate. Part (b) shows the output as a function of electrical discharge power. The gas mixtures ($He/N_2/CO_2$) in Torr are for curve A: 5/11/2; curve B: 5/7/3; curve C: 5/5/2; curve D: 5/7/2. (From W. B. Tiffany, R. Targ, and J. D. Foster, *Appl. Phys. Lett.* **15**, 91 (1969).)

in this fashion are called transverse flow CO_2 lasers. Models of multikilowatt carbon dioxide lasers, with continuous power up to 45 kW, have become available.

Models of slow axial flow, fast axial flow, and transverse flow CO_2 lasers are all available commercially, with somewhat overlapping power ranges. Table 3-4 summarizes some properties of the various types of CO_2 laser.

At relatively low power, the beam quality of the CO_2 laser can be excellent, but as output power increases, the beam divergence angle increases above the diffraction limit. Figure 3-10 shows a compilation of values of the beam divergence angle (in units of the diffraction limit) as a function of output power for a number of commercial CO_2 lasers. At outputs less than 500 W, most CO_2 lasers are near diffraction limited, but as the output climbs into the multikilowatt range, the beam divergence

Table 3-4 CO_2 Laser Types

Type	Scaling (W/m)	Gas velocity (m/sec)	Power (W)
Slow axial flow	60–80	<5	50–900
Fast axial flow	500	150	400–6,000
Transverse flow	1,000–2,000	60	5,000–45,000

angle increases markedly. This factor can limit the capability of a laser for industrial material processing applications, for which good focusing qualities are desired.

So far we have discussed basic CO_2 laser configurations. There have been a wide variety of variations of CO_2 laser types, for which we will briefly summarize a few properties.

TEA CO_2 lasers. A common variety of pulsed carbon dioxide laser is the so-called TEA (transversely excited, atmospheric pressure) laser. This is inherently a pulsed device rather than a continuous laser. In contrast to most CO_2 lasers, which operate at total gas pressures much less than 1 atm, the TEA laser operates at 1 atm gas pressure. This allows extraction of relatively large amounts of energy per pulse. But at these pressures, the uniform electrical discharge tends to transform into an arc discharge. To avoid this undesirable effect, the devices are operated in a relatively short pulse regime. The prime attraction of the TEA laser is its ability to

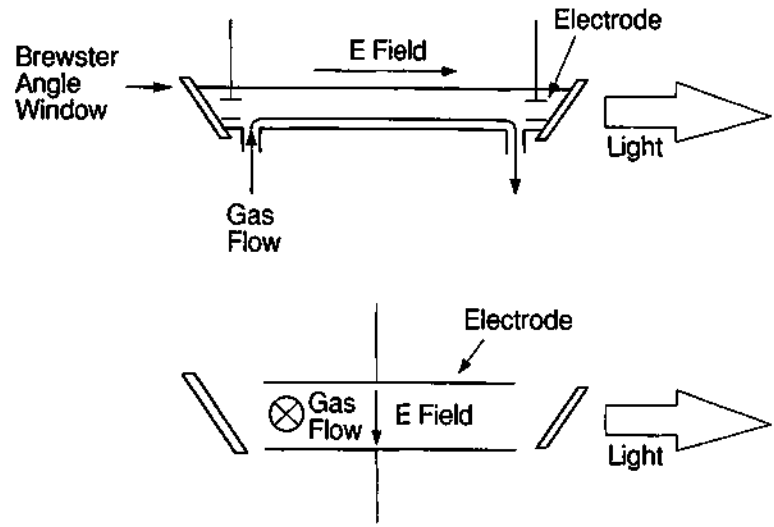


Figure 3-9 Evolution of CO_2 laser construction. The top portion shows the original type of construction, with axial gas flow and axial electrical discharge. The bottom portion shows the configuration with transverse gas flow and electrical discharge.

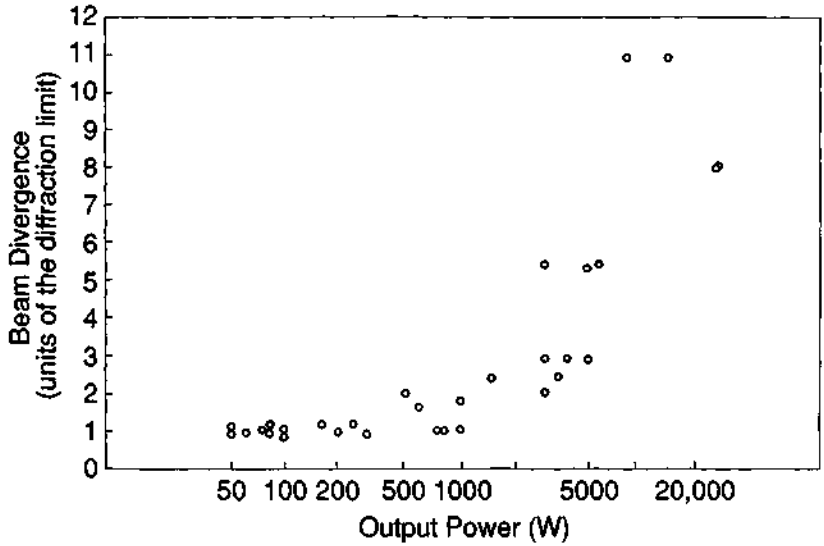


Figure 3-10 Beam divergence versus output power for selected commercial CO_2 lasers.

produce short intense pulses from a device that is modest in size. The TEA laser represents an important class of high-power pulsed infrared laser, with pulse durations in the submicrosecond to microsecond range and with peak powers in the multimegawatt regime. TEA lasers have important applications in industrial materials processing.

Waveguide CO_2 lasers. In these devices, hollow dielectric waveguides are employed both to confine the gas discharge and to guide the laser radiation. Compact, continuous lasers are available with high beam quality and with output powers of tens of watts. Waveguide CO_2 lasers are much smaller than conventional CO_2 lasers. They represent an important type of laser source for the relatively low power range and are widely used in applications for which their power output is adequate.

Gas dynamic lasers. In a gas dynamic laser, the population inversion is produced by expansion of the gas mixture through a nozzle, rather than by an electrical discharge. The gas is heated to high temperature and allowed to expand through a supersonic nozzle. As the temperature and pressure of the gas drops, the molecules in the upper laser level relax more slowly than molecules in the lower laser level, and a population inversion is produced some distance downstream from the nozzle. The optical cavity is positioned at this location, perpendicular to the gas flow. Gas dynamic lasers have been demonstrated with output power in excess of 100 kW. But after a number of years of research, disadvantages of gas dynamic lasers became apparent. Compared with electrical discharge lasers, gas dynamic lasers are larger, more complex, and less efficient. They have not been used extensively in industry.

Electron-beam-controlled lasers. These devices use a high-energy (100–200 keV) electron beam to ionize the gas. This allows separate optimization of the two functions of the electrons, sustaining the discharge and exciting the gas molecules. Such devices are scalable to very large pulsed outputs. They have been utilized in research for military applications, but they have not been widely employed in industrial usage.

We conclude the discussion of CO_2 lasers by summarizing some representative properties of available models. Table 3-5 presents properties of continuous CO_2 lasers, in three different ranges of power. Low power is defined somewhat arbitrarily as the range below 100 W; intermediate power as 100–3000 W; and high power as greater than 3000 W. Table 3-6 presents additional characteristics of the lasers in the intermediate power range, which covers many important industrial applications, in more detail. Finally, Table 3-7 presents characteristics of pulsed CO_2 lasers.

In summary, CO_2 lasers are the lasers that produce the highest levels of continuous power. They may be operated in a wide variety of different forms, with different methods of excitation. They form a very versatile class of lasers for industrial applications.

Other Molecular Lasers

There are a few other types of molecular lasers that have reached commercial status, and a larger number that have been demonstrated experimentally. Perhaps the most important of these lasers are the nitrogen laser and the carbon monoxide laser.

Table 3-5 **Continuous CO_2 Lasers**

Low power (3–100 W)
Mostly sealed tube or waveguide types
Single transverse mode
Sensing, scientific, and medical applications
Intermediate power (100–3000 W)
Slow axial flow models to 900 W
Fast axial flow models >400 W
Single mode (maybe donut) or "near" single mode
Typically diffraction-limited up to a few hundred watts and two times diffraction-limited at 1000 W
Applications primarily in material processing
High power (>3000 W)
Fast axial flow models to 6000 W
Transverse flow models >5000 W
Usually multimode
Applications in material processing and military uses

Table 3-6 **Intermediate-Power Continuous CO_2 Lasers**

Power: 100–3000 W
Beam diameter: 6–20 mm typical
Stability: ±2% typical
Programmability: Constant, ramps, steps, pulses (Peak pulse power ≈ 3 times continuous level)
Power distribution: Near Gaussian or donut below 1 KW
Beam divergence: 1–3 mrad typical
Spectrum: Several lines separated by about 0.06 μm
Scaling: 50–60 W/m for slow flow devices, 500 W/m for fast flow
Gas velocity: <5 m/sec for slow flow devices, 150 m/sec for fast flow
Gas consumption: 2–20 Standard Cubic Feet/Hour
Floor space: 30–60 ft^2

Table 3-7 **Pulsed CO_2 Lasers**

Output: Millijoules to 100 J
Pulse duration: 0.1–10 μsec; initial spike followed by lower-power tail is typical, duration tends to increase with output energy
Spatial profile: Mostly multimode
Beam divergence: 3–10 mrad typical
Flow: Mostly transverse, but all types available
Repetition rates: To kilohertz for selected models
Typical applications: Sensing, material processing, scientific, military

The nitrogen laser operating at 0.337 μm is a pulsed laser emitting pulses of nanosecond duration and peak powers up to megawatts. It is notable for its high-power pulses at an ultraviolet wavelength from a reasonably compact package. The nitrogen laser was developed early in the history of lasers, but problems with stability and beam profile have kept it from achieving widespread application. These problems seem to have been eased somewhat by recent developments.

The carbon monoxide laser operates simultaneously on a large number of wavelengths between 5 and 6 μm. There was once considerable interest in carbon monoxide lasers, because they offer relatively high efficiency and scalability to high power. Compared with carbon dioxide lasers, they offer a shorter wavelength, which would make them easier to use in applications such as cutting and welding of metals. A few such applications have in fact been demonstrated. But they have proven to be difficult to construct and to maintain, and in recent years, interest in carbon monoxide lasers appears to have waned.

A few other molecular lasers have limited availability. These include the nitrous oxide laser (10–11 μm), the water vapor laser (118 μm), and the hydrogen cyanide laser (311 and 337 μm). We note that the wavelengths can extend to very large values. At 337 μm, the wavelength is more than one-third of a millimeter. This is

remarkable in that it represents a laser with an almost macroscopic wavelength. Such long-wavelength lasers can be important sources for some specialized purposes.

4. *EXCIMER LASERS*

Excimer lasers are notable for their ability to produce high-power radiation in the ultraviolet portion of the spectrum. Excimer lasers are in principle scalable to large high-efficiency devices. Operation in the ultraviolet means that the diffraction-limited focal spot can be very small, smaller than for other high-power lasers. In addition, the short wavelength generally means that there will be good coupling of the energy to a workpiece.

The term "excimer" is derived from the word "dimer," meaning a diatomic molecule formed by the union of two atoms. If the molecule is in an excited state, it is referred to as an excited dimer, or excimer. The excimer lasers utilize molecules containing the noble gases, which do not form chemical compounds under normal circumstances. However, the noble gases may form compounds that have no stable ground state, but which may have excited states that are temporarily bound. An example is krypton fluoride. A gas mixture containing krypton and fluorine is excited in a pulsed electrical discharge. In a chain of complex processes, the metastable excited state KrF^* is produced. The asterisk denotes that the molecule is in an excited state. The excited state is bound for a short time and then dissociates according to the reaction

$$KrF^* \rightarrow Kr + F + h\nu \tag{3.1}$$

where $h\nu$ represents a photon with energy corresponding to a wavelength of 249 nm. Because there is no stable ground state, the population inversion needed for laser operation is easily obtained. Excimer lasers are necessarily pulsed devices, with pulse duration in the nanosecond regime.

The emission of the photon causes the KrF^* molecule to fall to its lowest energy state, in which the two atoms repel one another, so that the molecule breaks up. Thus, even though the noble gases like krypton do not form stable chemical compounds, it is possible for them to combine briefly in excited states to form transient short-lived molecules, which can emit laser light as they break up.

Excimer lasers have utilized two main methods of excitation: excitation by pulsed electric discharges or excitation by high-energy electron beams. Development of excimer lasers has branched into two channels, representing the two excitation methods. Electron-beam-excited devices are capable of producing very high energy pulses. Electron beam excitation involves large, expensive sources of high-energy electrons. Such devices can be scaled to very large size and are capable of reasonably high efficiency, potentially in the 5 to 10 percent range. Devices have been constructed with energy in the kilojoule range, and amplifiers with energy extraction capability in the hundred kilojoule range appear possible. Such devices are large and expensive and consume many resources. They have been used in areas such as laser-assisted thermonuclear fusion and military research.

Electric-discharge excimer lasers may be much smaller and less expensive. They are the devices that have found applications in industry for semiconductor fabrication, remote sensing, photochemistry, and material processing. Their energy extraction capabilities are much lower than those of the electron beam devices. Typical characteristics for commercial models are a pulse energy of a few tenths of a joule to a few joules per pulse and pulse repetition rates of tens to hundreds of hertz, with average power in the range of 100 W. Table 3-8 lists excimer laser types, wavelengths, and characteristics of commercially available excimer lasers. For a given device, the output is generally higher when krypton fluoride is used, rather than one of the other excimers.

The gas mixture for excimer formation contains appropriate noble gases (argon, krypton, xenon) and a halogen (fluorine, chlorine). The gas mixture is usually diluted with a carrier gas, like helium. The presence of the halogen gas involves safety issues, and the gas source is often isolated in an enclosed cabinet.

Excimer lasers have not been as reliable as some of the other types of lasers. Issues involving reliability include life of the gas mixture before it degrades and has to be replaced, corrosiveness of the gas mixture, life of electrodes, and life of the optics at the short wavelengths. Another difficulty with excimer laser use has been the rectangular shape of the output beam and the frequent presence of dark striations in it. This makes it more difficult to focus the beam to a uniform spot. These issues are being addressed by the laser manufacturers, and improvements have been forthcoming, so that excimer lasers have become useful tools for industrial use.

5. *METAL VAPOR LASERS*

Pulsed metal vapor lasers were first developed around 1966 but were slow to mature. The technology of these lasers has been difficult, primarily because of the high

Table 3-8 **Characteristics of Excimer Lasers**

Excimer	Wavelength (nm)
ArF	191
KrF	249
XeCl	308
XeF	351

Typical Characteristics for Commercial Devices

Pulse energy: 0.1–2 J
Pulse repetition rate: 10–500 Hz
Average power: To 150 W
Pulse duration: 10–30 nsec
Beam size: 10 × 30 mm
Beam divergence: 1–3 mrad

temperature at which the laser tube must be operated to keep the metal in vapor form at a reasonable pressure (around 1500°C in the case of gold). Because these lasers offer desirable wavelengths, they have been developed to commercial status and are now reliable, robust commercial products.

A metal vapor laser may consist of a ceramic tube with pellets of metal (such as gold or copper) positioned inside. The tube is surrounded by a cooling water jacket. An electrical discharge through a gas (neon) in the tube heats the metal and produces a low-pressure vapor. The laser is essentially a pulsed device with a high pulse repetition rate (kilohertz), so that the beam appears continuous to the eye. The beam diameters are typically 1 in. or more, larger than those of most familiar visible lasers.

Commercial metal vapor lasers are copper (511 and 578 nm) and gold (628 nm). Experimental demonstrations have included lead (723 nm), manganese (534 nm), and barium (1500 nm). The availability of these wavelengths from small devices with short pulse duration has allowed development of a number of novel applications, such as photodynamic therapy in medicine and high-speed photography. Although such lasers are still rather uncommon, they are beginning to be utilized for industrial material processing. The short wavelength and high peak power allow high irradiance to be delivered to a small area on a workpiece.

Commercial copper vapor lasers emit average powers ranging from a few watts to more than 100 W. The two wavelengths 511 and 578 nm are emitted simultaneously, with about two-thirds of the power at the shorter wavelength. The output is in the form of short pulses with duration in the 10–40 nsec range at pulse repetition rates up to 20 KHz. The input power required ranges from a few kilowatts for small devices to around 16 kW for larger ones. Typical small units operate on 220 VAC single-phase input, and larger models use 220 or 415 V three-phase input. Commercial pulsed gold vapor lasers can emit up to several watts of average output power. The other characteristics are similar to those of the copper vapor lasers.

Of the metal vapor lasers, copper is the most highly developed. Table 3-9 presents some characteristics of commercial metal vapor lasers.

Table 3-9 **Characteristics of Metal Vapor Lasers**

Commercial wavelengths: 511 and 578 nm (copper) and 628 nm (gold)
Average power: To 100 W (copper). 20 W typical
Pulse Duration: Tens of microseconds
Pulse repetition rates: 5–10 kHz
Beam divergence: 4–8 mrad
Beam diameter: 25–50 mm
Typical applications: Medical, photobiology, forensic, flow visualization, high-speed photography

B. Solid State Lasers

Solid state lasers are characterized by active media involving ions of an element present in a small percentage in a solid host material. The laser material is usually in the form of a cylindrical rod with the ends polished flat and parallel, although sometimes slabs or disks are used. The pumping is accomplished by optical excitation. A schematic diagram of a typical lamp-pumped solid state laser is shown in Figure 3-11. The lamp and laser rod are positioned side by side, inside a reflecting structure that allows a large fraction of the pumping light to reach the laser rod. The Q-switch is a device that can shorten the duration of the pulse and increase its power. Q-switches are discussed in Chapter 5.

The prototype of the solid state laser is the ruby laser, the first laser ever operated, but the most common type has become Nd-doped yttrium aluminum garnet (YAG). The term "solid state" is used in a sense different from its common usage. In connection with lasers, it does not imply the use of semiconductors; rather, it applies to lasers involving impurity ions in insulating host materials.

The ions that are usually employed are those of the transition metals, like chromium or cobalt, or of the rare earth elements, that is, the group of elements with atomic numbers from 58 to 71. These elements share the common feature of an interior unfilled shell of electrons, which leads to a narrow fluorescent linewidth. According to Equation (1.9), narrow fluorescent linewidth is favorable for laser operation, because it leads to high gain and to reduced requirements for the minimum population inversion necessary for laser operation.

The host materials are hard, gemlike crystalline materials, or alternatively, glasses. Such materials can readily be fabricated into rods with polished ends.

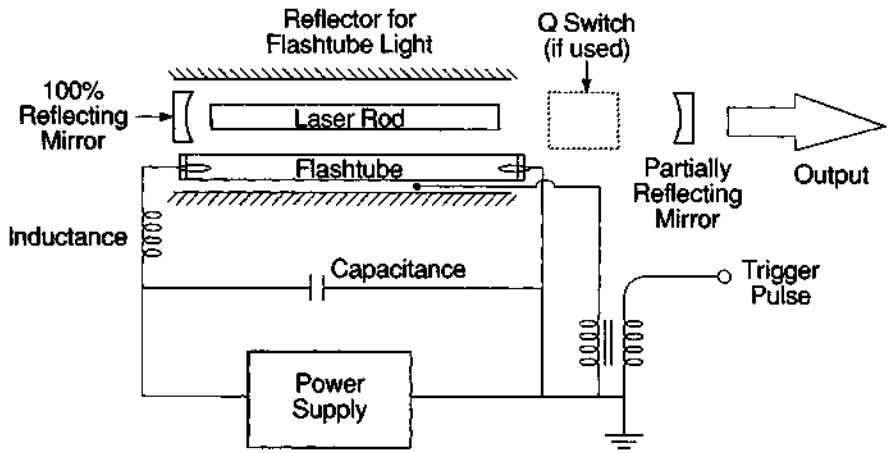


Figure 3-11 Structure of a typical lamp-pumped solid state laser.

Table 3-10 presents a listing of commercially available solid state lasers. The table groups the lasers according to the commonality of their use. The first column lists the laser material, giving first the chemical element that provides the energy levels used for laser operation and then, after a colon, the host material. The materials listed with more than one element before the host use the additional elements for energy transfer to increase efficiency. The second (and third) listed elements (frequently chromium) absorb the pump energy and then transfer it resonantly to the first listed element, which provides the laser action. The gemlike hosts are identified at the bottom of the table.

The second column lists the wavelengths at which continuous operation is available, and the third column lists the wavelengths at which pulsed operation is available. We note that the list of commercially available lasers changes frequently, with new types being added. In addition to commercially available solid state lasers, many other types have been demonstrated in laboratory operation.

The first three groupings represent lasers pumped by flashlamps or arc lamps. The fourth category presents the lasers that are available as semiconductor-diode-pumped models.

In the table, the neodymium-based lasers usually operate at a wavelength near 1.06 μm: the shorter wavelengths are provided by frequency doubling, tripling, and quadrupling using nonlinear optical effects. (See Chapter 5 for a description of the nonlinear optical effects that lead to generation of light at frequencies that are integer multiples of the fundamental laser frequency.) The wavelengths near 1.32 μm arise from a different set of energy levels. We shall describe a few of the most common types of these lasers in more detail.

1. NEODYMIUM:YAG LASERS

The neodymium-doped yttrium aluminum garnet laser, denoted Nd:YAG, has been by far the most commonly employed solid state laser for many years. It is available in a wide variety of forms, including continuous, normal pulsed, and Q-switched. The use of nonlinear optical elements allows generation of a number of wavelengths shorter than the 1.06 μm fundamental wavelength. The host material, YAG, has the chemical formula $Y_3Al_5O_{12}$. The Nd resides in the crystalline lattice in substitutional sites for yttrium as a trivalent ion, with its outer electrons removed. It typically is present at a level around 2 percent.

YAG offers a number of desirable properties. It is a hard, stable, isotropic material. Crystal growth techniques yield YAG material of high optical quality, and YAG can be polished to a good optical finish. It has high thermal conductivity, about ten times that of glass, allowing for relatively easy extraction of the heat generated. The high thermal conductivity is an essential feature. It allows dissipation of heat at high thermal loading, which in turn allows for operation at high values of average output power. Although many other crystalline hosts for neodymium have been developed, YAG has remained as the most widely used. Only yttrium lithium fluoride (YLF) appears to offer significant competition.

Table 3-10 **Solid State Lasers**

Very common: Many models available

Laser	Continuous wavelengths (μm)	Pulsed wavelengths (μm)
Nd:YAG	1.06, 0.266, 0.355, 0.532, 1.32	1.06, 0.266, 0.355, 0.532, 1.32

Reasonably common: Several models available

Laser	Continuous wavelengths (μm)	Pulsed wavelengths (μm)
Ruby	—	0.694
Nd:glass	—	1.06, 0.266, 0.355, 0.532
Alexandrite	—	tunable 0.72–0.78
Ti:sapphire	tunable 0.7–1.1	tunable 0.7–1.1
Nd:YLF	1.047, 1.31	1.047, 1.31, 0.351, 0.523
Ho:YAG	2.1	2.1
Er:YAG	2.94	2.94

Uncommon: Typically only one model

Laser	Continuous wavelengths (μm)	Pulsed wavelengths (μm)
Co:MgF_2	—	tunable 1.75–2.5
Cr:Th:Ho:YAG	—	2.09
Er:glass	—	1.54
Er:Cr:YSGG	—	2.79
Er:YSGG	—	2.79
Ho:YSGG	—	2.1
Nd:Cr:GSGG	—	1.06
Tm:YAG	—	2.01

Available as diode-pumped lasers

Laser	Continuous wavelengths (μm)	Pulsed wavelengths (μm)
Nd:YAG	1.06, 0.532	1.06, 0.532, 1.32
Nd:YLF	1.047, 1.32	1.047, 0.523, 1.32
Nd:Cr:GSGG	1.06	—
Er:Yb:glass	1.535	—

Host abbreviations: YAG, yttrium aluminum garnet; YLF yttrium lithium fluoride; YSGG, yttrium scandium gallium garnet; GSGG, gadolinium scandium gallium garnet.

The energy levels of the neodymium ion relevant to laser operation are shown in Figure 3-12. The energy levels are characteristic of the neodymium ion, with small shifts with host material. See, for example, the wavelengths of Nd:YAG and Nd:YLF in Table 3-10. The figure shows the spectroscopic designations of the

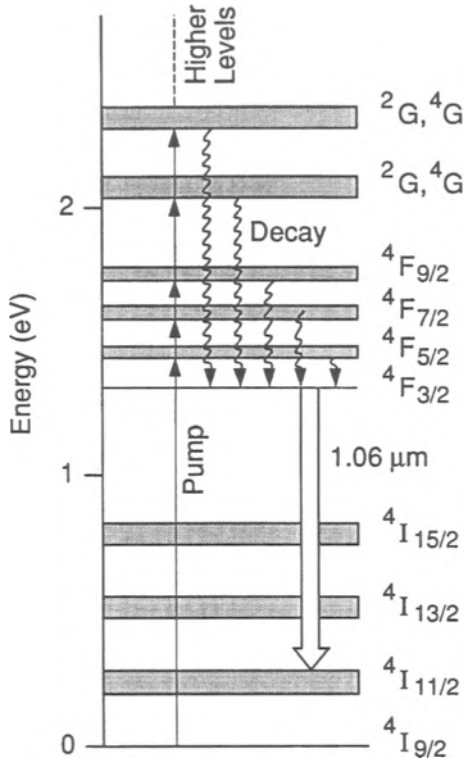


Figure 3-12 Energy levels of the Nd ion. The spectroscopic notation for the levels is indicated. The excitation path and laser transition are shown.

energy levels. The pumping from a flashlamp is into a number of levels that lie above 12,000 wavenumbers. These levels all relax nonradiatively to the $^4F_{3/2}$ level. The laser transition is between the $^4F_{3/2}$ level and the $^4I_{11/2}$ level, which lies about 2000 wavenumbers above the ground level. Thus, this is a four-level laser system. Because the $^4I_{11/2}$ level will be unpopulated at ambient temperature, it is easy to obtain a population inversion with modest amounts of pumping energy. It is not only the low threshold that makes this host material of particular importance; there are other hosts that provide comparably low threshold. Rather, it is a combination of favorable properties, including good optical quality, high thermal conductivity, stability, and narrow fluorescent linewidth, that have made YAG a desirable host material.

The output characteristics of Nd:YAG lasers may conveniently be discussed under three categories: continuous, pulsed, and continuously pumped, repetitively Q-switched. Table 3-11 summarizes the properties of available Nd:YAG lasers in these three modes of operation.

In continuous operation, the laser is excited by continuous arc lamps or, increasingly, by semiconductor diode lasers. If arc lamps are used, the emission from

Table 3-11 **Nd:YAG Laser Properties**

Lamp-pumped operation at 1064 nm
Continuous operation: TEM_{00} operation to 35 W, many models in 10–20 W range; multimode operation to 2400 W, many models in 50–100 W range.
Normal pulse operation: Pulse energy to 100 J; typical pulse repetition rates of a few hertz, but rates to 300 Hz at reduced energy; pulse durations in 0.2–20 msec range
Q-switched operation: Pulse energy to 2 J at low repetition rate; kilohertz repetition rates at reduced energy, typical pulse duration in 7–10 nsec range
Lamp-pumped frequency-doubled operation
Q-switched operation: Pulse energy to 1 J at low repetition rate at 532 nm; repetition rates to 50 kHz, but at reduced pulse energy; pulse durations typically 5–10 nsec; at 532 nm, pulse energy typically 40–50 percent of that at 1064 nm; some models also operate at 355 and 266 nm with further reduced energy per pulse
Diode-pumped operation at 1064 nm
Continuous operation: Outputs to 30 W multimode
Normal pulse operation: Outputs to 30 mJ multimode; repetition rates to 1 kHz; pulse durations in few hundred microsecond range
Q-switched operation: Outputs to 200 mJ; many models in range of tens of microjoules, repetition rates to 50 kHz at reduced energy; pulse durations around 10 nsec
Diode-pumped frequency-doubled operation
Continuous operation: To 5 W TEM_{00}
Normal pulse: 1 mJ pulses at 10 Hz with 200 μsec duration
Q-switched operation: 10–60 μJ/pulse; repetition rates to kilohertz; typically 10 nsec pulse duration

krypton-filled lamps gives a better match to the absorption spectrum of Nd:YAG than xenon-filled lamps; thus, krypton lamps are generally employed, especially at higher levels of power.

If pulses with relatively large pulse energy are desired, the Nd:YAG is excited by a flashlamp. This yields pulses at relatively low pulse repetition rate. One may trade off the pulse energy, pulse duration, pulse repetition rate, and average power for such lasers. For example, as pulse energy increases, the maximum pulse repetition rate usually decreases.

A third type of operation is the continuously pumped, repetitively Q-switched mode. The pump source (arc lamp or semiconductor laser) remains on continuously. The laser is Q-switched, usually by an acoustooptic Q-switch (see Chapter 5). The pulse repetition rate can be very high (kilohertz), but the pulse energy is relatively low, typically millijoules. Again, there are trade-offs between pulse energy, pulse duration, pulse repetition rate, and average power. Figure 3-13 shows how the peak power and average power vary with pulse repetition rate, for a commercial device.

At low pulse repetition rate, the peak power is independent of pulse rate, and so the average power increases with pulse rate. It can reach levels close to 100 percent of what would be available in continuous operation with the same laser. At pulse rates above 1 kHz or so, the peak power decreases with increasing pulse rate, so that the average power does not continue to increase. Such lasers may operate at repetition rates approaching 100 kHz, but at those high rates, the pulse energy is low and the pulse duration increases compared with its value at lower rates.

Another variation of Nd:YAG laser operation that has become widely available commercially involves frequency doubling. The output of the frequency-doubled Nd:YAG laser is in the green portion of the visible spectrum at 532 nm. The technique of frequency doubling is described in Chapter 5 in the section on Nonlinear Optical Elements. Most frequency-doubled Nd:YAG lasers are operated in a nanosecond pulse regime, because the efficiency of doubling increases as the peak power becomes higher. But some continuous models are available. Frequency-doubled Nd:YAG lasers are one of the most important sources of high laser power operating in the visible spectrum. They are often used in applications where small focal spot size is desired, such as in micromachining. Some properties of frequency-doubled Nd:YAG lasers are included in Table 3-11.

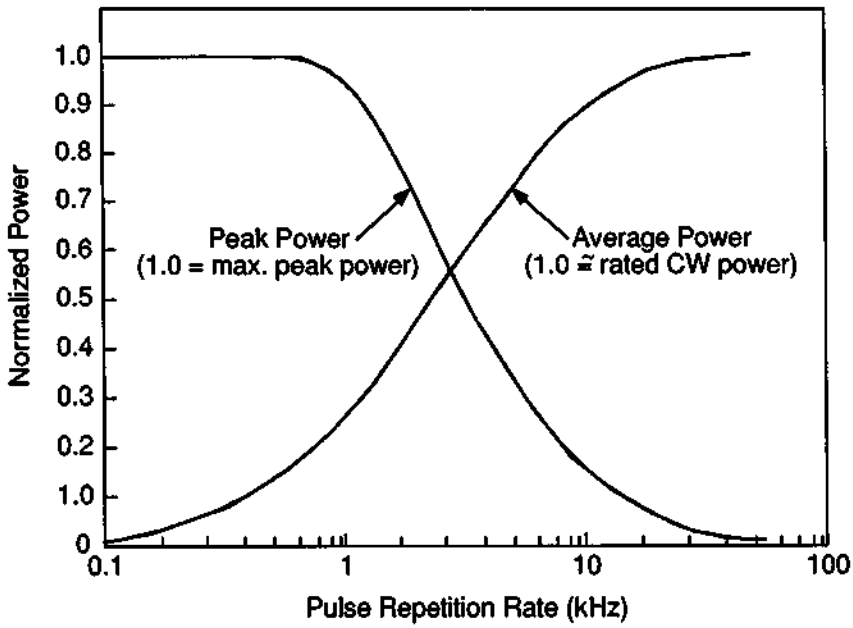


Figure 3-13 Peak power and average power as a function of pulse repetition rate for a continuously pumped, repetitively *Q*-switched Nd:YAG laser. (Based on information from Excel Quantronix.)

In addition to frequency-doubled operation, the Nd:YAG laser is available in frequency-tripled (355 nm) and frequency-quadrupled (266 nm) versions.

Nd:YAG lasers have sometimes been constructed with the material in a form other than a cylindrical rod. One variation is the slab laser. The material is a slab, with the ends cut at Brewster's angle. The beam bounces back and forth between the top and the bottom of the slab, confined by total internal reflection. The mirrors that define the laser cavity are external to the slab. Slab lasers offer the desirable feature of having a relatively large volume of material, so as to be scalable to high power.

So far, we have discussed Nd:YAG lasers pumped by arc lamps and by flashlamps. The use of semiconductor laser diodes as the pump source for solid state lasers has become common in the 1990s, and a variety of commercial diode-pumped lasers has become available. The properties of semiconductor lasers will be discussed in the next section.

Semiconductor lasers offer relatively high efficiency of conversion of electrical input to optical output. In addition, the wavelength of the semiconductor laser may be matched to a strong absorption line of Nd:YAG, so that the efficiency of the Nd:YAG laser system becomes higher than is possible with arc lamp or flashlamp pumping. Figure 3-14 shows the absorption spectrum of Nd:YAG and the emission spectra of some pump sources. A flashlamp may have a broad continuum emission, like a blackbody source, so that much of its energy falls in regions where the Nd:YAG does not absorb. A Kr arc lamp has line emission in the regions where Nd:YAG absorbs, and thus it is more efficient for excitation of Nd:YAG. The output of the semiconductor laser diode is adjusted (by composition and temperature) to be near 810 nm, near the peak of the neodymium absorption. Thus, almost all of the semiconductor laser energy may be absorbed. The diode lasers are themselves relatively efficient, and their output is absorbed better by the Nd:YAG than the light from flashlamps or arc lamps. Thus, diode-pumped solid state lasers have much higher efficiency than conventionally pumped devices. There is much less waste heat to remove. For equal output power, a diode-pumped Nd:YAG laser requires much less cooling than a lamp-pumped model. All these factors combine to reduce the size and resource requirements of diode-pumped devices. Diode-pumped solid state lasers are much more compact and efficient than earlier solid state lasers. They are still in a state of rapid development in the mid 1990s, with their capabilities increasing rapidly. Some of their properties are also included in Table 3-11.

One configuration for a diode-pumped solid state laser is shown in Figure 3-15. This arrangement is referred to as an end-pumped configuration, because the light from the diode is incident on the end of the Nd:YAG crystal. A potassium titanyl phosphate (KTP) frequency doubling crystal is included inside the laser cavity. Both mirrors have high reflection at 1064 nm so that the Nd:YAG laser light will build up to a high level inside the cavity. The rear mirror has high transmission near 810 nm to allow the diode laser light to enter the Nd:YAG. The output mirror has high transmission at 532 nm to allow the green frequency-doubled light to emerge. The arrangement shown here is efficient because it allows the diode light to be focused

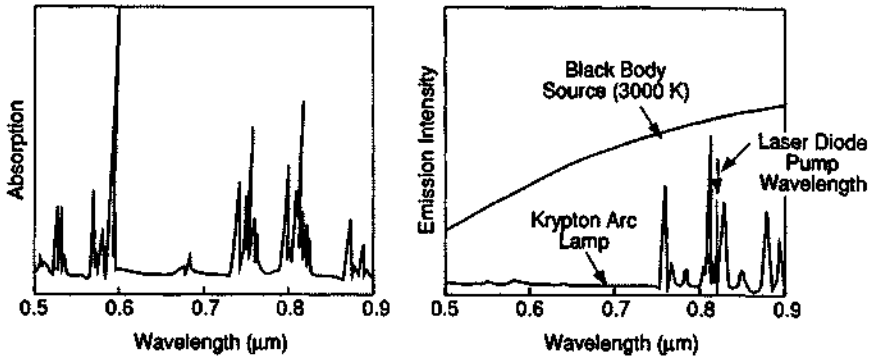


Figure 3-14 Left: Absorption spectrum of Nd:YAG. Right: Emission spectra of some pump sources.

so as to match the volume of the Nd:YAG in which the Gaussian mode is generated. But it does not allow the use of larger arrays of diode lasers. To use more pump lasers and increase the diode laser pumping power, one may use side pumping, a configuration in which many diode lasers are directed at the side of the Nd:YAG. Alternatively, the diode lasers may be located remotely and their light brought to the Nd:YAG through the use of optical fibers.

Many models of diode-pumped Nd:YAG lasers have become available. Their output power does not yet match the maximum power available from lamp-pumped models, but because of their advantages of compactness, efficiency, and lowered power requirements, they are replacing lamp-pumped devices for many applications.

2. *NEODYMIUM:GLASS LASERS*

Neodymium-doped glass is another important solid state laser material. Glass has a number of desirable characteristics as a laser host. It can be produced in large pieces of high optical quality and can be fabricated into a variety of sizes and shapes, ranging from fibers with diameters of a few micrometers to rods more than 1 m long to slabs and disks with sizes of tens of centimeters in each of two dimensions. The thermal conductivity of glass is lower than that of most crystalline hosts. Thus, for a Nd:glass laser to operate at high average power or at high pulse repetition rate, the glass laser must have at least one small dimension in order to allow rapid heat removal. A second important difference between glass and crystalline host materials is that the emission lines of ions in glass are broader. This raises the threshold for laser operation, because a larger population inversion is required to achieve the same gain, according to Equation (1.11). On the other hand, the broader absorption lines in glass lead to increased absorption of the pump light and allow greater energy storage.

The broader emission lines also allow mode-locking over a broader spectral range than is possible with other materials. Nd:glass lasers thus offer properties

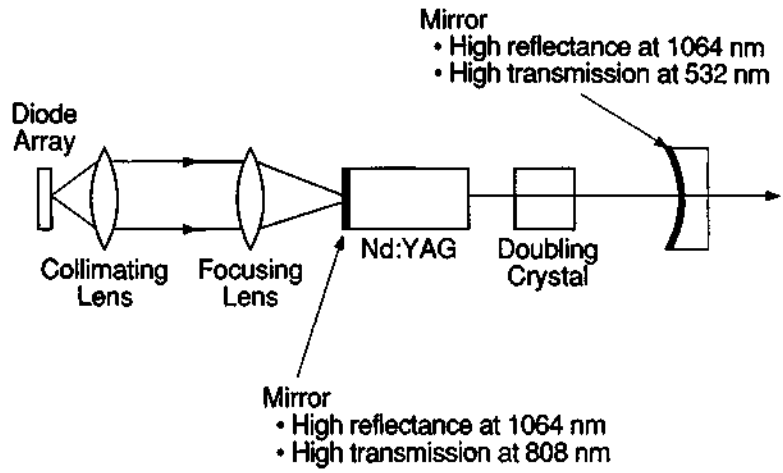


Figure 3-15 End-pumped configuration for a diode-laser-pumped intracavity frequency-doubled Nd:YAG laser.

favoring the production of picosecond pulses. For applications requiring high energy outputs, high radiance, or high peak power, glass systems generally outperform crystalline hosts.

Other differences between Nd:glass and Nd:YAG lasers arise from differences in the thermal conductivity, in the fluorescent lifetime, and in the ability to fabricate large pieces of the materials. The longer fluorescent lifetime, of Nd:glass and also the ability to dope to high concentration allow for greater energy storage per unit volume. That glass may be fabricated into very large pieces, much larger than is possible with YAG, was just discussed. Thus, glass is the material of choice for large laser systems where it is desired to extract the largest amount of energy.

But the higher thermal conductivity of YAG allows greater capabilities for high average power. This includes both continuous operation and operation at high pulse repetition rates. Thus, both materials are employed in practical laser systems for different types of applications.

Neodymium has been incorporated into a wide variety of glasses, including silicate-based glasses, phosphate-based glasses, and fluorophosphate-based glasses. However a glass with an alkali–alkaline earth base appears to offer the best combination of relatively long fluorescent lifetime, high fluorescent efficiency, and good durability. For example, a rubidium barium silicate glass has a relatively long fluorescent lifetime of 900 μsec, a property desirable for large energy storage per unit volume.

Three typical regimes of operation may be identified for Nd:glass lasers. These are the normal pulse mode, with pulse duration around 1 to 10 msec, Q-switched operation with pulse duration of some tens of nanoseconds, and picosecond pulse operation. The duration of the normal pulse operation may be varied by varying the duration of the pumping light pulse.

The fluorescent lifetime of typical glasses is a function of the doping level of neodymium. Typically, the fluorescent lifetime starts to decrease as the neodymium concentration increases above 2 percent. For pulse duration longer than 1 msec, repumping by the pump source is required. The normal pulse Nd:glass laser exhibits relaxation oscillation behavior, with microsecond duration substructure, as was described in Chapter 2. Normal pulse Nd:glass lasers are available with pulse energy up to hundreds of joules.

Q-switched Nd:glass lasers are available with pulse duration in the nanosecond regime, and with pulse energy ranging from a fraction of one joule to more than 100 J. In the higher-energy range, the laser is usually configured as a master oscillator, power amplifier system. The pulse is generated in a relatively small laser, the master oscillator, and then is amplified by passage through a number of larger rods (or slabs or disks) of Nd:glass. The amplifiers are optically pumped so as to produce a population inversion. They have no end mirrors and cannot operate as lasers by themselves. When the pulse from the oscillator travels through an amplifier, the energy stored in the population is extracted by stimulated emission in a single pass. Thus, the pulse energy is increased without increasing the pulse duration. The size of the amplifier is increased in each stage so as to keep the power per unit area low enough to avoid damage. This helps reduce the problem of optical damage. See Chapter 6.

To obtain picosecond pulses, one may select one of a train of mode-locked pulses from a Q-switched laser. Because of the broad fluorescent linewidth of Nd:glass, the spectral width across which mode-locking may occur is of the order of 10^{12} Hz. The mode-locked pulse may then have a duration around 10^{-12} sec. Such devices are also usually constructed in a master oscillator, power amplifier configuration. Commercially available devices can generate 1.5 J of energy in a 1.5 picosecond (psec) duration pulse, for a peak power around 1 terawatt (TW).

Nd:glass lasers have sometimes been configured as slab lasers, which were also discussed with reference to Nd:YAG. They offer a relatively large volume of material and high values of energy storage. Some commercial models of Nd:YAG slab lasers are available.

The very large lasers used for thermonuclear fusion research have usually been Nd:glass, often frequency-doubled or -tripled. The Nova laser at Lawrence Livermore National Laboratory, for example, used a total path of 6 m of Nd:glass in a master oscillator, power amplifier configuration, and it was capable of generating more than 10 KJ of energy at 1.05 μm wavelength in a 2.5 nsec duration pulse and more than 1 KJ in a 100 psec pulse [1].

3. *Nd:YLF Lasers*

The neodymium ion has been incorporated into a very wide variety of host materials, and it is not appropriate to review them all here. But one host material that has become important recently is yttrium lithium fluoride ($YLiF_4$), commonly referred to as YLF. Nd:YLF lasers have become increasingly available.

Nd:YLF has lower thermal conductivity than Nd:YAG, but its thermal conductivity is high enough to allow it to operate continuously at room temperature. In addition, it reduces one problem present in Nd:YAG, thermal lensing. Optical pumping of Nd:YAG leads to a radial temperature gradient, which makes the laser rod act like a thick positive lens. Nd:YLF is less subject to this thermal lensing effect. As a result, the beam quality may be improved. In one comparative study of the performance of neodymium in various hosts [2], the best pulsed beam quality was obtained with Nd:YLF.

Nd:YLF is a birefringent material. As a result, it can operate at two different wavelengths, 1.047 and 1.053 μm, depending on the polarization. In commercial models, the shorter wavelength appears to be more common, although both wavelengths are available. Nd:YLF lasers are available as continuous and as continuously pumped, repetitively Q-switched devices. Frequency-doubled and frequency-tripled models also are available. There are a number of models that incorporate pumping by semiconductor laser diodes.

The Nd:YLF laser may be considered as a viable competitor to Nd:YAG for some applications, especially when pulsed beam quality is important.

4. RUBY LASERS

Ruby is the material in which laser operation was first demonstrated, in 1960. Ruby is sapphire (crystalline aluminum oxide) in which a small percentage (typically 0.05 percent) of the aluminum has been replaced by chromium. The chromium resides in the crystalline lattice of the sapphire as a trivalent ion. The energy level diagram of ruby has already been shown in Figure 1-9. Ruby absorbs light in the green and blue portions of the visible spectrum and emits red fluorescence at 0.6943 μm, which is also the wavelength of laser operation. Ruby is a three-level system, so that more than 50 percent of the electrons must be excited to cause a population inversion. Thus, the laser threshold of ruby is relatively high, and intense optical pumping is required.

Ruby has high thermal conductivity, and heat may be dissipated well. But the high threshold and intense pumping generate so much heat that ruby lasers are not capable of operation at high values of average power. Experimental continuous ruby lasers have been demonstrated, but these have not progressed beyond laboratory devices. For practical purposes, ruby lasers are pulsed devices, with low pulse repetition rates, usually less than one pulse per second.

Although ruby lasers were widely used in the beginning of the laser era, they seem less common today. Although there are a number of commercial models available, there are far fewer than for Nd:YAG. They are important because many of the early studies that defined the capabilities and technology of lasers were performed with them.

One function in which ruby lasers still excel is their ability to generate pulses of very high peak power from a relatively simple, compact package. A Q-switched ruby laser with a peak power of hundreds of megawatts is relatively easy to develop.

But at levels in excess of a few hundred megawatts, the high optical power may damage the ruby crystal. See Chapter 6.

Ruby lasers are available as normal pulse lasers, with pulse duration in the millisecond regime. Pulse energies range up to 400 J, although models with output around 10 J are more common. Q-switched models with pulse duration of some tens of nanoseconds are also available, with peak power ranging up to 1000 MW.

5. *HOLMIUM LASERS*

Lasers based on the rare earth element holmium are becoming fairly readily available. They operate at a wavelength near 2.1 μm, depending on the host material. The most common host is YAG, with several models of Ho:YAG lasers available. Holmium laser materials commonly incorporate additional rare earth elements (such as erbium and thulium) as sensitizers. The sensitizers absorb the pumping energy more effectively than holmium. They then transfer the energy to the holmium ions via lattice interactions. Holmium lasers have been operated both with lamp pumping and semiconductor laser pumping.

Although holmium lasers are not nearly as common as Nd:YAG lasers, they are potentially important for some applications, including remote sensing in the atmosphere and medical applications. They offer the advantage that their light is absorbed in the front structures of the eye and does not reach the retina. Thus, they are less hazardous to the eye than Nd:YAG.

6. *VIBRONIC SOLID STATE LASERS*

A relatively new class of solid state lasers that offers tunability of the output wavelength is the so-called vibronic solid state laser. In these materials, the active material is an ion of a transition metal (like Cr, Co, or Ti) present as a substitutional impurity in a crystalline lattice. In a narrow-linewidth laser like Nd:YAG, only the electronic state changes. In the vibronic lasers, both the electronic state and the vibrational state change. This is illustrated in Figure 3-16. The horizontal axis is the distance between ions in the lattice. This distance may vary because of lattice vibrations. The energy of the lattice vibrations is quantized. The quantum of lattice energy is called a phonon. The electronic states are represented by the parabolic curves, which represent the potential wells in which the ions reside. The horizontal lines in the wells are the different vibrational states. The scalloped lines near the edges of the wells indicate phonon transitions between the vibrational levels. The wavelength of both absorption and emission vary because there is a range of different energies over which the transitions may occur. The effect is to broaden the absorption and emission lines, and to make the absorption occur at a somewhat different wavelength than the emission. This situation is similar to that of the dye laser, which will be described later. These devices are also called tunable phonon-terminated solid state lasers, because, as Figure 3-16 shows, there may be emission of several phonons after the emission occurs.

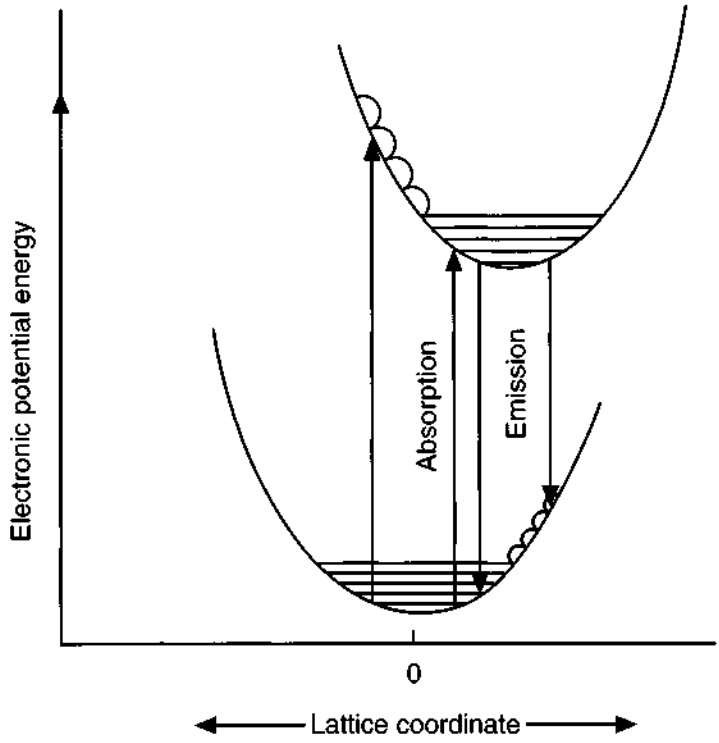


Figure 3-16 Energy level structure for a vibronic solid state laser.

The two leading types of these lasers are alexandrite and Ti:sapphire. Alexandrite is chromium-doped beryllium aluminate ($Cr:BeAl_2O_4$). Alexandrite lasers are tunable over the range from about 710 to 820 nm, although most commercial models do not cover this entire range. These lasers are usually pulsed, pumped by flashlamps, but continuous models are available. Continuous alexandrite lasers are usually pumped by arc lamps, but they have been pumped with semiconductor laser diodes. Alexandrite lasers have the unusual property that their operating characteristics improve when they are heated above room temperature. A number of commercial alexandrite lasers are available, including continuous, normal pulse, and Q-switched models.

Titanium-doped sapphire (Ti:sapphire) has excellent characteristics for a tunable laser. Ti:sapphire is a four-level system and has high thermal conductivity. It has a very broad tuning range, 670–1050 nm. When Ti:sapphire lasers are frequency-doubled, tuning over the range 335–525 nm is added. They may be operated either continuously or pulsed. The peak of the absorption is around 500 nm, so that it is often pumped by another laser, such as argon or doubled Nd:YAG. Because its absorption is not in the range where semiconductor lasers operate, it is not suitable for semiconduc-tor laser pumping. Many commercial models are available, including continuous, normal pulse, and Q-switched models.

Ti:sapphire may be operated mode-locked and generate very short pulses, as short as 50 femtoseconds (fsec). Because of their very broad tunability, they are displacing dye lasers for spectroscopic applications, especially in the red end of the visible spectrum and in the near infrared.

Two other vibronic solid state lasers are current research favorites and appear to be developing rapidly. These are based on chromium-doped LiSAF and LiCAF (lithium strontium aluminum fluoride and lithium calcium aluminum fluoride). Cr:LiSAF, for example, has an absorption peak near 670 nm and has tunable emission over the range 840–960 nm. Both of these laser materials can be pumped by semiconductor lasers, and they could in principle be developed as small all-solid-state packages. These lasers are still in an early stage of development, and it is too soon to judge their eventual potential, but they show promise.

C. Semiconductor Lasers

Semiconductor lasers (also called diode lasers or injection lasers) employ structures that are much different from the gas and solid state lasers described earlier. Semiconductor lasers were first operated in 1962 and thus are almost as old as the other types of lasers. They represent an entirely different approach to laser fabrication. They provide lasers with properties different from those of the lasers that we have discussed.

A semiconductor laser uses a small chip of semiconducting material as the active medium. In size and appearance, the semiconductor laser is similar to a transistor or to a semiconductor light emitting diode. The most common semiconductor laser material is the alloy aluminum gallium arsenide. We will return later to a discussion of the materials commonly employed for semiconductor lasers.

1. SEMICONDUCTORS AND BAND STRUCTURE

We begin by briefly reviewing some basic properties of semiconductors. Electrical conductivity in semiconductors occurs because of the presence of a relatively small number of free charge carriers that can move through the crystalline lattice of the material under the action of an applied electric field. If the carriers are electrons, the semiconductor is called an *n*-type semiconductor. If the conductivity occurs because of the motion of holes (a hole is a missing electron in one of the pairs of bonding electrons between atoms, and it acts like a positive charge), the material is called *p*-type. The junction between *p*- and *n*-type material is called a *p–n* junction, and it acts as a diode. Such diodes have been used in many solid state electronic applications, such as rectifiers and light emitting diodes (LEDs). A *p–n* junction is also the region in which laser action occurs in a semiconductor laser.

Semiconductor LEDs are nonlaser devices. They have been used in many display applications. They use materials similar to those used in lasers, and they rely on the presence of a *p–n* junction, as do semiconductor lasers. However, they have radiance lower than laser diodes and are a device distinct from lasers.

A solid state laser, like Nd:YAG, involves an isolated impurity ion embedded in a crystalline host lattice. The radiative transitions that produce laser operation occur between discrete electronic energy levels of the impurity. The host lattice serves mainly to maintain the ions at fixed positions. In contrast, in the semiconductor laser, the relevant energy levels are characteristic of the entire crystalline lattice. These are not discrete levels, but multitudinous levels that lie so close together as to form a continuum, commonly called an energy band. There are two energy bands of interest for a semiconductor laser, the valence band and the conduction band. The valence band is the highest-lying band filled with electrons. The conduction band lies higher in energy and is separated from the valence band by the so-called energy bandgap, a region in which there are no allowed electronic states. Because of the presence of the bandgap and because the energy levels below the gap are filled, it is difficult to accelerate electrons with an applied electric field. Hence, the electrical conductivity of the material is low. Because a few electrons may be present in the conduction band, because of thermal effects or doping of the materials, and these electrons may be accelerated by an electric field, there is some small amount of electrical conductivity. Hence, the materials are called semiconductors.

When light with photon energy greater than the energy of the bandgap is incident on the material, an electron may be raised from the valence band to the conduction band. This leaves an absent electron, or a hole, in the conduction band. Both the hole and electron are free to move through the material, with the hole acting like a positive charge carrier. Thus, the electrical conductivity is increased.

An electron in the conduction band may fall back across the bandgap and recombine with a hole in the valence band. The difference in energy is emitted as light.

The energy band diagram near a p–n junction is shown in Figure 3-17. The material is made n– or p–type by doping with suitable impurities, so that there are free electrons in the conduction band in the n–type material and holes in the valence band in the p–type material. For example, if a pentavalent impurity like antimony is added to aluminum gallium arsenide, that has an average of four valence electrons per atom, there will be an extra electron that is not needed for bonding. This electron will be free to move and will contribute to n–type conductivity. Similarly, a trivalent impurity like indium means that there is a missing bonding electron (a hole), which will contribute to p–type conductivity.

The so-called Fermi level, denoted E_f, is the energy to which the electronic states are filled. In the neighborhood of the junction, the energy bands undergo a shift as indicated. There is an energy barrier that restricts electrons in the conduction band from flowing to the right, or holes in the valence band from flowing to the left. Thus, the junction has electrical rectification properties.

The left portion of Figure 3-17 shows the energy band structure with no voltage applied. Electrons left of the junction and holes to the right of the junction are separated by the barrier. The right portion of the figure shows how the structure changes when a forward bias voltage is applied to the junction. The barrier height is reduced. Some of the electrons in the conduction band will overlap some of the holes in the valence band, as indicated. There are filled electronic states lying above empty

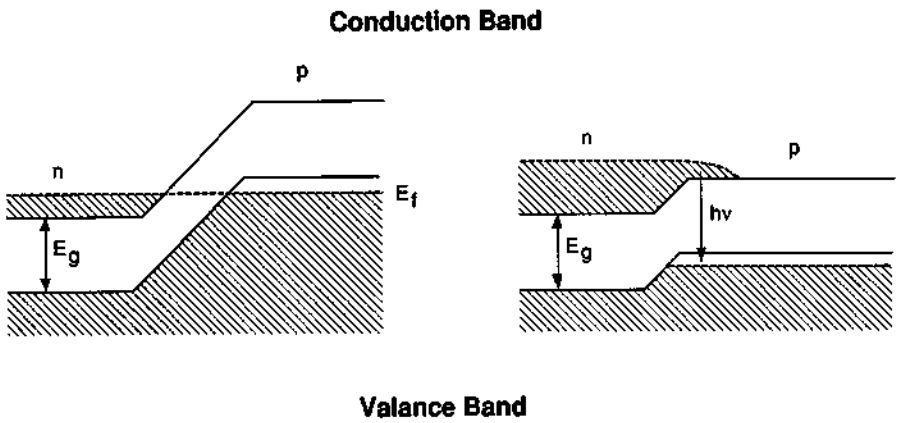


Figure 3-17 Band structure near a p–n junction in a semiconductor. E_g denotes the bandgap energy. Left: With no bias voltage. Right: With a forward bias voltage.

states. Thus, the applied voltage has created a population inversion, the necessary condition for laser operation.

Radiative recombination of the holes and electrons can occur. Electrons fall across the energy gap and fill the empty states in the valence band. The difference in energy is emitted as light. This is indicated in the figure as photons of energy $h\nu$. If the population inversion is large enough, stimulated emission can occur and the junction region can act as a laser. In operation, electrons are injected into the device from the n-type side; thus, these lasers are frequently called injection lasers.

The thickness of the junction region is small, typically of the order of 1 μm. Light traveling in the plane of the junction is amplified more than light perpendicular to the junction. The laser emission emerges parallel to the plane of the junction.

2. *SEMICONDUCTOR LASER PROPERTIES*

We begin the discussion of semiconductor laser structure with a rectangular parallelepiped as illustrated in Figure 3-18. The dimensions often are a few hundred micrometers, up to perhaps 1 mm. The junction forms a plane within the structure. Two of the sides perpendicular to the junction may purposely be roughened so as to reduce their reflectivity. The other two sides are optically flat and parallel, and these form the mirrors for the laser. They may be prepared either by cleaving or polishing. The reflectivity of the air–semiconductor interface is high enough (often around 36 percent) that no other mirrors are needed, Sometimes, one reflecting surface may be coated to high reflectivity in order to decrease losses and enhance laser operation. The laser light is emitted perpendicular to these surfaces, as indicated. The arrangements for electrical contacts and for heat sinking are also indicated.

The structure shown in Figure 3-18 is the first type of semiconductor laser structure developed. It consists simply of a junction between p- and n-type layers of the

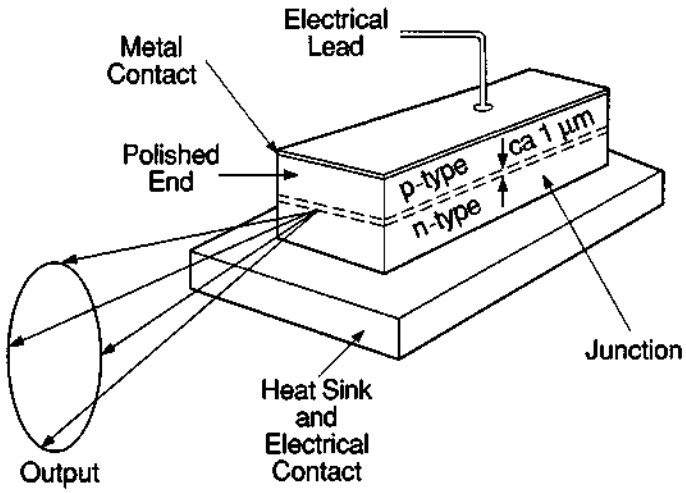


Figure 3-18 A simple semiconductor laser structure.

same material. This structure now is obsolete. Modern semiconductor lasers contain many layers of material of varying composition, as we shall describe later.

For laser operation to occur, the population inversion must be high enough so that optical gain exceeds loss. The current through the junction must provide enough holes and electrons so that the radiation generated by their recombination exceeds the losses. There are a number of loss mechanisms in semiconductor lasers. They include spreading of light out of the active region, absorption of light by free carriers in the junction region, and transmission of light through the end surfaces. Thus, there is a threshold value of the electrical current for laser operation to occur.

Figure 3-19 shows how the output of a semiconductor laser varies with the input current. At low values of the current, the device is a light emitting diode (LED). It emits a relatively small amount of incoherent light. The LED region has a relatively slow increase of light output with current. At the laser threshold, where the population inversion is large enough so that gain by stimulated emission can overcome the losses, the slope of the curve changes. As current increases above the threshold, the light output increases much more rapidly than in the LED region. The light is now coherent laser light.

A parameter called slope efficiency characterizes the laser output. The slope efficiency is the increase in optical output power divided by the increase in electrical input power. The definition applies only to the region where the output is increasing linearly. Slope efficiency is an important specification, which is used to characterize the laser performance. Commercial semiconductor lasers with slope efficiencies around 30 percent are available.

If the light output is too high, the semiconductor material may be damaged. The figure shows the existence of a damage threshold at the right end of the linear slope.

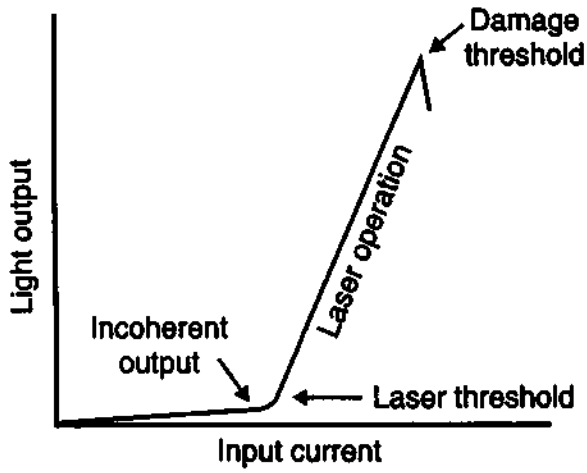


Figure 3-19 Typical diode laser output versus input current.

At the damage threshold, catastrophic optical damage (COD) begins. Above the threshold, permanent damage occurs and the light output drops rapidly with increasing current. We will discuss COD more fully in Chapter 6. For now, we simply note the existence of a damage threshold, which should not be exceeded. The manufacturer of laser diodes usually will specify a maximum current.

Ideally, the light output should increase linearly with current, as shown in Figure 3-19. In some cases, the curve of output power as a function of input current may show kinks or bends. These are regions in which the light output does not increase linearly with input current. This is generally an undesirable situation, which reduces the stability of the device and which may disqualify it from many applications The concept of slope efficiency is not valid if kinks are present. The manufacturers of laser diodes have devoted extensive effort to produce stable devices that do not exhibit this type of behavior.

The output of a semiconductor laser diode often has multiple longitudinal modes present. Because the length of the laser cavity is very short, typically less than 1 mm, the spacing of the modes is relatively large. The spectral characteristics of the output may often be of the form shown in Figure 3-20, which shows the longitudinal mode spectrum of a commercial semiconductor laser operating at a wavelength near 1550 nm. Many longitudinal modes are present, and the spectrum spreads over a breadth of almost 20 nm. Thus, the spectral width of the emission from a semiconductor laser may be much broader than that of a gas laser.

The emission spectrum may be narrowed to a single longitudinal mode, and the spectral width narrowed considerably, by a number of means. With a type of laser configuration called a gain-guided structure, to be defined later, the spectrum near threshold for laser operation may be as shown in Figure 3-20, but the spectrum narrows as the input current increases, and the number of longitudinal modes decreases. At currents well above threshold, there may be only one longitudinal mode present.

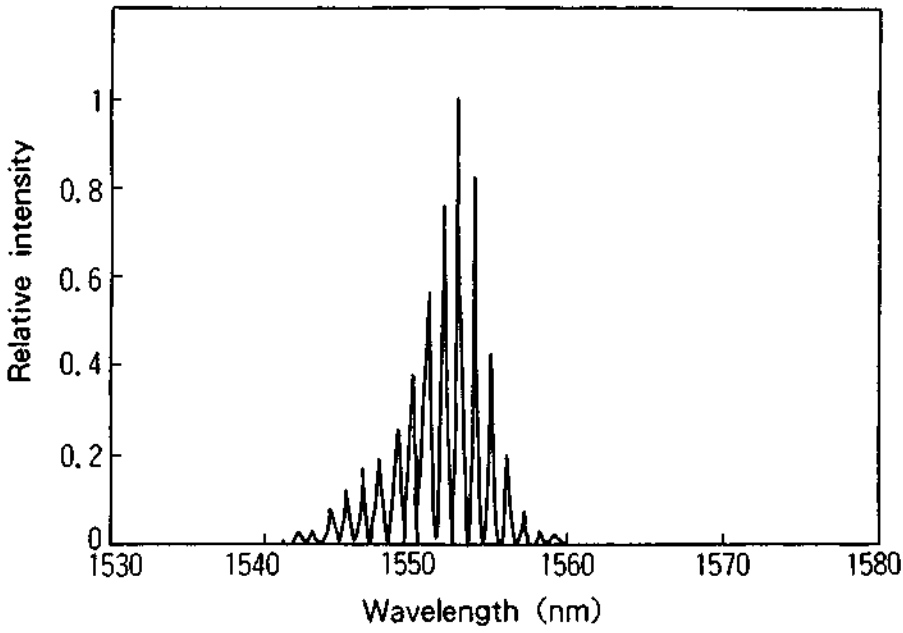


Figure 3-20 Spectral emission of a commercial semiconductor laser operating at a wavelength near 1550 nm, showing the presence of a number of longitudinal modes.

In other cases, an external cavity may be added to the laser to restrict the output to a single mode. Thus, there are options commercially available for which the output spectrum may be quite narrow, consisting of a single longitudinal mode.

The output characteristics of diode lasers change with the operating temperature. Figure 3-21 shows how the output changes with temperature for typical semiconductor lasers. As temperature increases, the threshold current increases. At a given value of current above threshold, the output decreases as the temperature becomes higher. We note that semiconductor lasers may be operated above room temperature. The laser manufacturer often will specify a maximum operating temperature.

The threshold current has an exponential dependence on temperature, and is given by

$$I_t(T_2) = I_t(T_1) \times \exp\left[\frac{T_2 - T_1}{T_0}\right] \tag{3.2}$$

where $I_t(T_2)$ and $I_t(T_1)$ are the threshold currents at temperatures T_2 and T_1, respectively, and T_0 is a scale factor, which typically has a value around 150 K. As an example, if T_1 = 0°C and T_2 = 25°C, the threshold current will be about 18 percent higher at the higher temperature than at the lower.

The exact wavelength of the emission from a semiconductor laser depends on several factors, including the composition of the semiconductor material, the operating temperature, and the current. The wavelength of the laser is set first by the

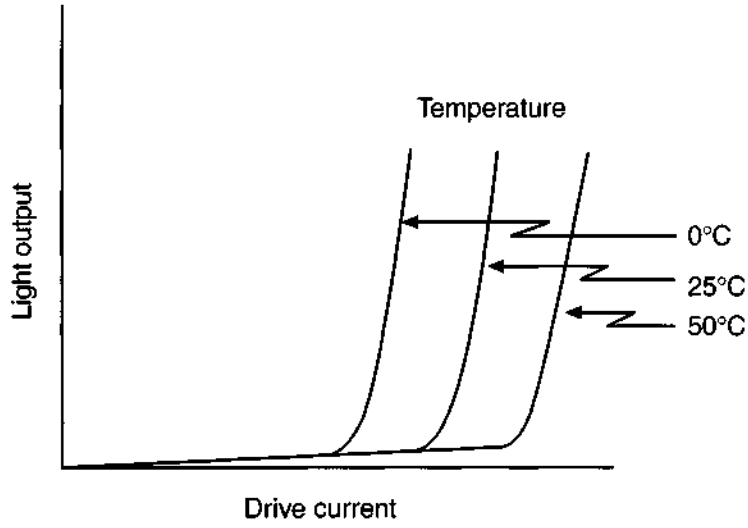


Figure 3-21 Output of a semiconductor laser as a function of input for various operating temperatures.

composition of the semiconductor material, to an approximate value. Then, it may be tuned over some relatively small range by changing the operating temperature and the current. Figure 3-22 shows temperature tuning of a commercial semiconductor laser operating in a single longitudinal mode and at a wavelength near 810 nm. The specified temperature is the temperature of the diode laser case. Further fine-tuning is possible by varying the input current. As the current increases, the temperature within the case increases slightly and the laser output moves to a slightly longer wavelength. The behavior shown in the figure, with tuning over a range of 10 nm or so, may be regarded as typical.

The figure also shows the presence of mode hopping, a common characteristic for lasers that operate in a single longitudinal mode. The mode hopping is characterized by abrupt jumps in the wavelength of operation as the temperature changes. This arises because the laser can shift from operation in one longitudinal mode to a different longitudinal mode. The result is a sudden shift in operating wavelength. Mode hopping can be a problem if one is trying to adjust the laser operation to a specified wavelength. Mode hopping can be eliminated by use of external stabilizing cavities, so that the temperature tuning curve will be smooth and continuous.

If the laser operates in multiple longitudinal modes, as was illustrated in Figure 3-20, the temperature tuning takes the form of a gradual shift of the entire mode spectrum to longer wavelengths as temperature increases.

The spatial profile of the beam emitted from a semiconductor laser is usually not round, as in the case of familiar gas lasers. The laser emission often has an elliptical spatial profile, as illustrated in Figure 3-23. This phenomenon is caused by diffraction. In the direction perpendicular to the junction, the beam is confined by the nar-

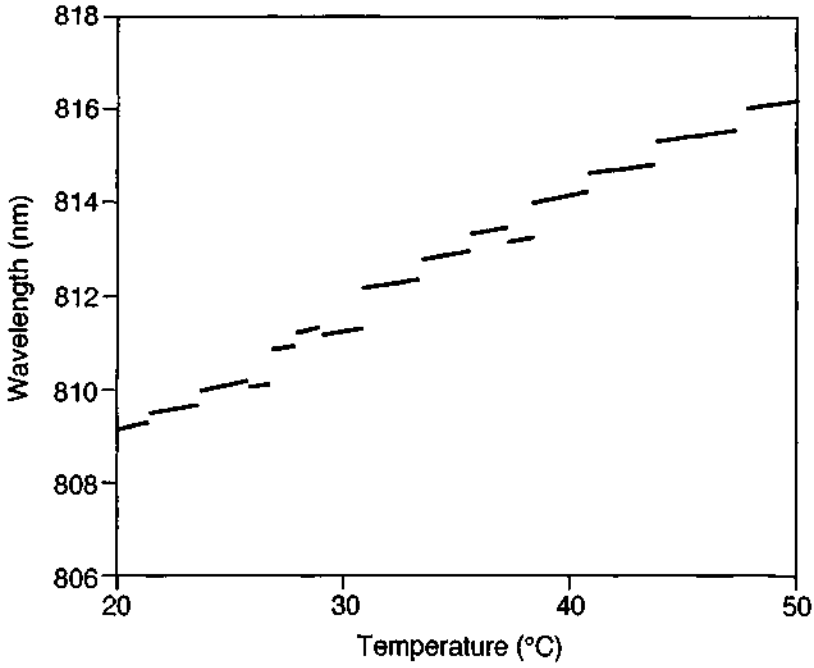


Figure 3-22 Variation of the wavelength of operation for a commercial semiconductor laser operating at a wavelength near 810 nm with case temperature. The discontinuous jumps are due to mode hopping.

row junction, which often is around 1 μm wide. The beam in that direction may spread to an angle as large as several tens of degrees. In the direction parallel to the junction, the beam is not confined so tightly and spreads less, often to around 10° or so. The result is a fan-shaped beam, as indicated in the figure.

A laser beam with this shape is said to be astigmatic. The astigmatism arises because the apparent sources of the emission for the radiation parallel and perpendicular to the junction are separated. The numerical value of the astigmatism is a distance equal to the amount of the separation. A typical value for the astigmatism may be around 10 μm.

Often, one uses anamorphic lenses (lenses with different magnification in each of two directions perpendicular to the direction of light propagation) to correct the astigmatism, at least partially. This causes the beam to have a circular profile. But the beam divergence is still larger than for familiar gas lasers. Gas lasers, with a diffraction-limited beam emerging from an aperture with a diameter around 1.5 mm may have a circular beam with divergence of a few tenths of a degree. Because of their larger beam divergence, semiconductor lasers have lower radiance than other lasers. The larger beam divergence also means that the light from semiconductor lasers cannot be focused as well as the light from better-collimated lasers.

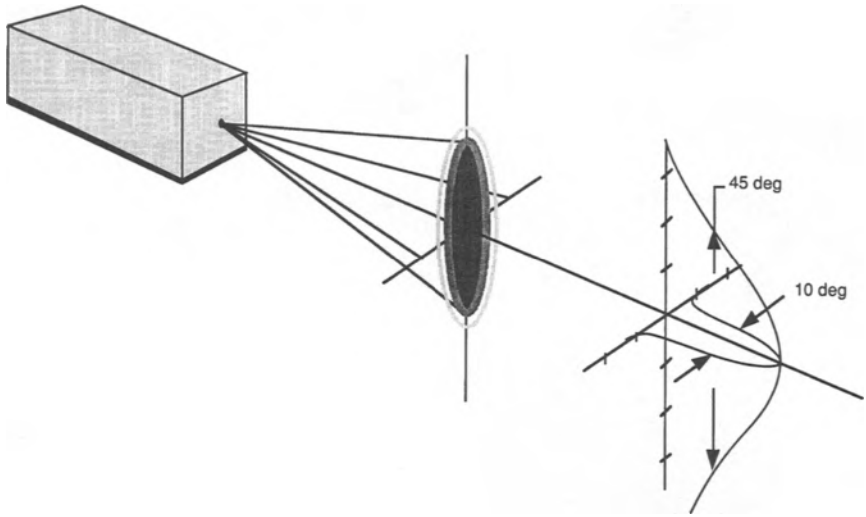


Figure 3-23 Typical beam profile from a semiconductor laser.

Semiconductor lasers may operate continuously (CW) at room temperature or even above. They may be pulsed by pulsing the input current, often with pulse durations in the nanosecond regime. They are also often operated in a so-called quasi-CW mode, which involves a relatively long pulse duration, typically of the order of 100 μsec. In quasi-CW operation, the peak power is higher than in continuous operation. Quasi-CW operation is well suited for the pumping of solid state laser materials.

Semiconductor lasers have many attractive features. These include relatively high efficiency, small size, light weight, low power consumption, and the capability to be driven by simple low-voltage power supplies. They are manufactured by the tens of millions each year. Major applications include use as light sources for compact disc players, printers, magnetooptic data storage, and optical fiber telecommunications. Visible semiconductor lasers operating in the red region of the spectrum are replacing helium–neon lasers in some applications.

3. *SEMICONDUCTOR LASER MATERIALS*

We now consider the materials used to fabricate semiconductor lasers. Most semiconductor lasers are based on III–V compounds, which are compounds formed by elements from column III and column V of the periodic table. Gallium arsenide, the first semiconductor laser material, is an example. The first semiconductor lasers were fabricated from crystals containing a junction between p- and n-type gallium arsenide. Such lasers are now obsolete and have been superseded by structures formed of alloys containing three or four elements from columns III and V of the periodic table. These alloys are called ternary or quaternary compound semiconductors, respectively. One example of a ternary compound is $Al_{1-x}Ga_xAs$ (aluminum

gallium arsenide). An example of a quaternary compound is $In_{1-x}Ga_xAs_{1-y}P_y$ (indium gallium arsenide phosphide). In this notation, the numbers x and y are composition parameters, which define the exact composition. They may have values ranging from 0 to 1. Thus, the $Al_{1-x}Ga_xAs$ system can vary continuously from AlAs ($x = 0$) to GaAs ($x = 1$).

The properties of the material also change as x and y vary, Figure 3-24 shows how the index of refraction varies in the $Al_{1-x}Ga_xAs$ system as the composition changes.

The most common of the semiconductor lasers is the $Al_{1-x}Ga_xAs$ laser. The output wavelength varies with the composition parameter x. Commercial devices are available with output wavelength ranging from about 780 to 880 nm. $Al_{1-x}Ga_xAs$ lasers are used for many applications including optical disks, compact disc players, and lightwave communications.

A second semiconductor laser material is the quaternary alloy $In_{1-x}Ga_xAs_{1-y}P_y$, which has two composition parameters, x and y. The values of x and y may be varied independently to adjust the wavelength of operation. They are also used to match the lattice constant of the substrate on which the material is grown. $In_{1-x}Ga_xAs_{1-y}P_y$ lasers are available with output wavelengths in the range from about 1150 to 1650 nm. They have been widely employed as sources for optical fiber telecommunications.

Another important type is $Al_xGa_yIn_{1-x-y}P$; here, Al, Ga, and In are all column III elements. This alloy is notable for operation in the red portion of the visible spectrum,

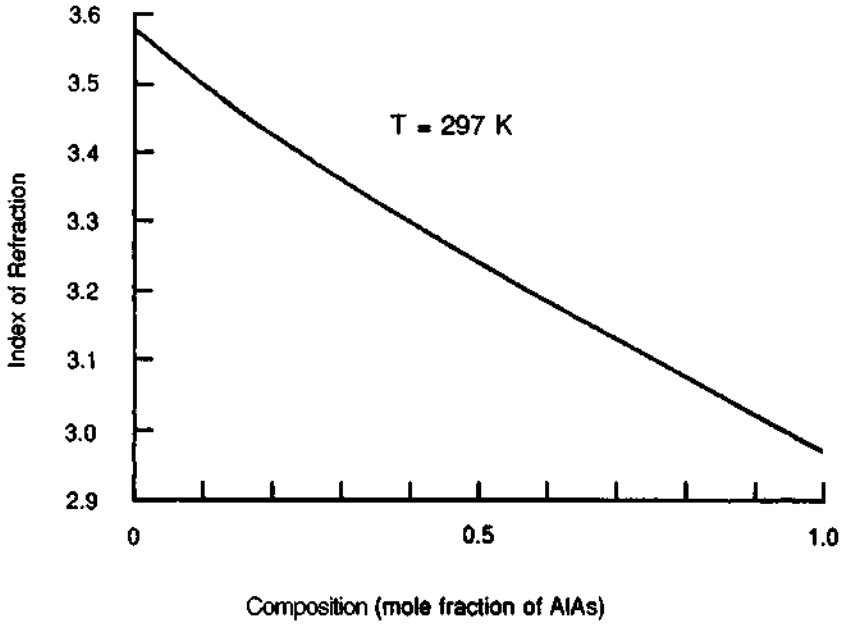


Figure 3-24 Variation of index of refraction with composition in the aluminum gallium arsenide system.

covering a wavelength range from 630 to 680 nm. Lasers based on it are beginning to replace helium–neon lasers in some of their applications.

There are also some commercial devices available for operation at 980 nm, a wavelength useful for excitation of erbium-doped fiber amplifiers (EDFA) for lightwave communications. EDFAs will be discussed in Chapter 24.

Table 3-12 summarizes some of the most important types of semiconductor laser materials. These lasers are all commercially available from a number of sources. Figure 3-25 illustrates commercial availability of some common semiconductor lasers in the mid 1990s. The dark areas show the ranges of wavelength and output power available. The development of these lasers has been driven largely by the needs of lightwave telecommunication. Devices with wavelength in the 830, 1300, and 1550 nm regions have been developed for this important application field. Many millions of these devices are manufactured and sold annually. The price of relatively low power (a few milliwatts) semiconductor lasers in the 800 nm region is low, perhaps five to ten dollars in quantity. The price increases rapidly as output power increases.

There are other semiconductor laser types available, based on II–VI compounds, such as $PbS_{1-x}Se_x$. These devices operate in the long wavelength infrared region. They are very costly and require cryogenic operation. They are far less common than the lasers we have described based on III–V compounds. They are available with outputs in the range from 3 to 25 μm. At the short-wavelength end of this range, they are available as standard products. At the longer wavelengths, they are available on special order. Table 3-13 presents some examples of commercially available devices. The range of operation for a particular compound is determined by the composition parameter x. For a specific value of x, the general wavelength region is determined. The wavelength of operation may then be tuned over a range by varying the temperature. If one varies both composition and temperature, the indicated range of wavelengths may be covered.

Table 3-12 **Semiconductor Laser Materials**

Material	Wavelength range (nm)
$Al_{1-x}Ga_xAs$	780–880
$In_{1-x}Ga_xAs_{1-y}P_y$	1150–1650
$Al_xGa_yIn_{1-x-y}P$	630–680
$In_{1-x}Ga_xAs$	980

Table 3-13 **Examples of Long-Wavelength Infrared Semiconductor Laser Materials**

Material	Wavelength range (μm)
$Pb_{1-x}Sn_xSe$	8–12.5
PbSe	8
$Pb_{1-x}Eu_xSe$	3.5–8

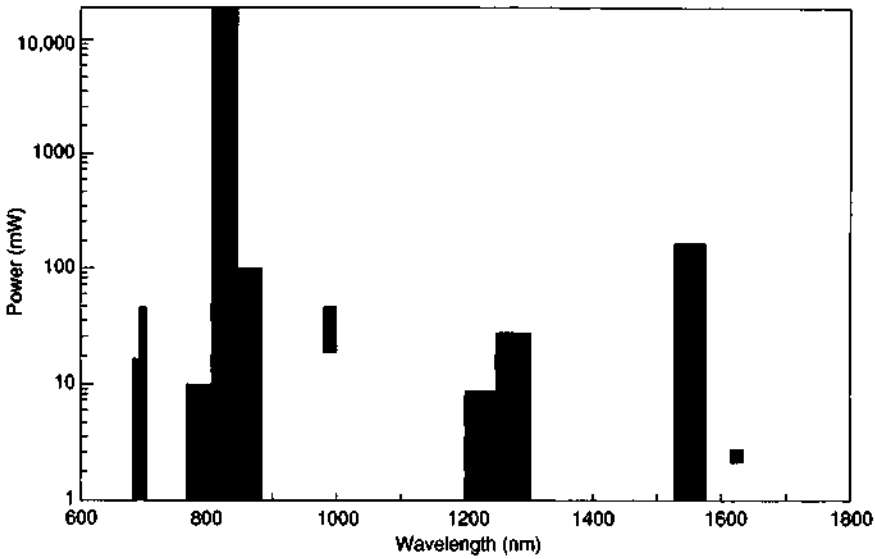


Figure 3-25 Commercial availability of continuous semiconductor diode lasers based on III–V compounds as functions of wavelength and output power.

4. DIODE STRUCTURES

We turn now to a discussion of the structure of a semiconductor laser diode. We previously introduced for purposes of discussion a simple and now obsolete structure, a rectangular block of material with a p-type region, an n-type region, and a p–n junction plane parallel to one of the surfaces, as indicated in Figure 3-18. Modern diode lasers are formed of more complex structures, which contain a number of thin layers of material of varying composition. One example, out of many possible, of the types of layers that are possible is shown in Figure 3-26. This figure illustrates the complexity of the structures that have been developed. The material is grown in thin layers, typically with thickness of the order of one or a few micrometers, on a substrate of semiconductor material. The growth is accomplished by carefully controlled epitaxial growth techniques, such as molecular beam epitaxy or vapor phase epitaxy. Such growth techniques allow deposition of very thin layers of material of specified composition as single crystalline layers with high perfection.

The structures typically contain several layers of material of varying composition and doping levels, each of which serves a specific purpose, such as confinement of light. We shall describe the functions of the regions of varying composition shortly. At some position, a p–n junction region is formed. The junction region is the region where laser operation occurs. The notation p^+ and n^+ in the figure refers to heavily doped conductive regions of p- and n-type material, respectively.

Figure 3-26 also indicates some other features, such as the use of proton-bombarded areas, which have low electrical conductivity. The proton-bombarded areas form a stripe that confines the current and defines the area where laser

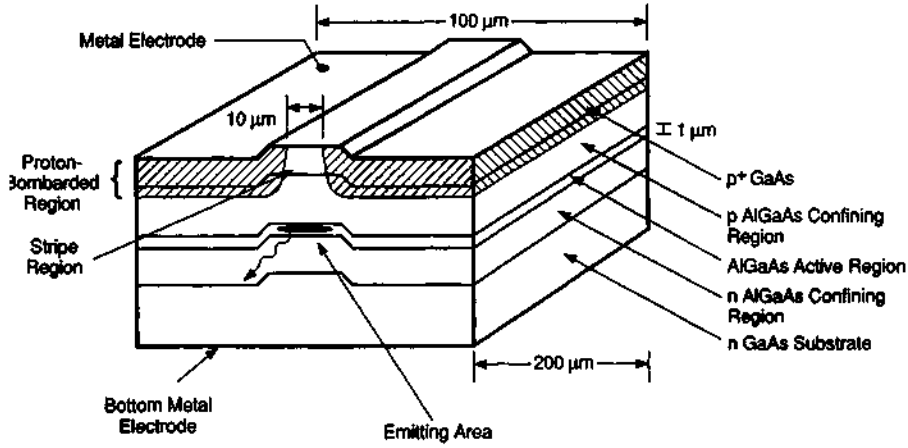


Figure 3-26 Example of a typical type of structure for a semiconductor laser diode, showing the different types of semiconductor layers commonly employed.

operation will occur. The incorporation of a stripe, either in this fashion or by selective metallization of the surface so that the electrical contact is in the form of a stripe, is a very common feature of semiconductor lasers. Figure 3-27 shows the incorporation of a stripe by means of selective insulator and metallization layers. The current flows only in the region where the metallization contacts the semiconductor. Many commercial lasers incorporate stripes in this fashion.

We now consider the functions of the different layers of semiconductor materials, as illustrated in Figures 3-26 and 3-27. The heavily doped layers are electrically conducting. They facilitate making the electrical contacts. The sequence of layers *n*-AlGaAs/AlGaAs/*p*-AlGaAs defines the light emitting region in the center layer. It forms what is called a double heterostructure, in which there are two changes of composition of the material as one goes through the active region. To understand the function of the double heterostructure, let us first consider the earliest devices, as illustrated in Figure 3-18. In this simple form, the junction was formed by a junction between *p*- and *n*-type GaAs. This was a material with a single composition; the junction was called a homojunction (also called a homostructure). Figure 3-28a

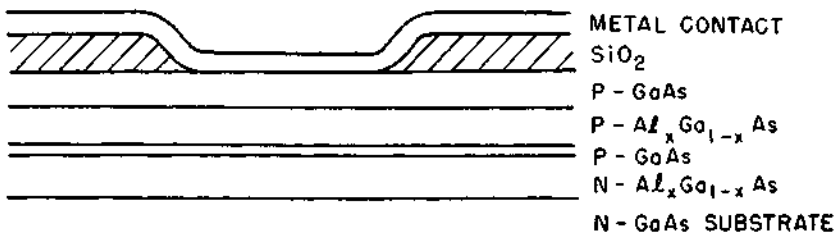


Figure 3-27 Example of a stripe geometry diode structure in which the stripe is formed by selective contacting of the metallization to the semiconductor.

shows a homojunction. It is called a homojunction because the material on each side of the junction has the same composition.

Figure 3-28 also shows how the index of refraction of the material changes through the junction. In part (a), variation in doping of the material provides a small step in the index of refraction, as indicated on the right side of the figure. This tends to provide some confinement of light in the region of the junction, because of total internal reflection. However, the step in the index of refraction is small and the confinement is not strong. Losses due to spreading of light out of the active region are large, so that the input currents are large. These devices were short-lived and subject to damage because of the high values of current required. Homojunction diode lasers are now extinct.

The use of a single heterojunction (also called a single heterostructure), as shown in Figure 3-28b, provided better confinement of light in the junction region. In this structure, there is one change in composition of the material through the junction, and so the device is called a single heterojunction. The structure provides a large change in index of refraction at one side of the active region, according to the data

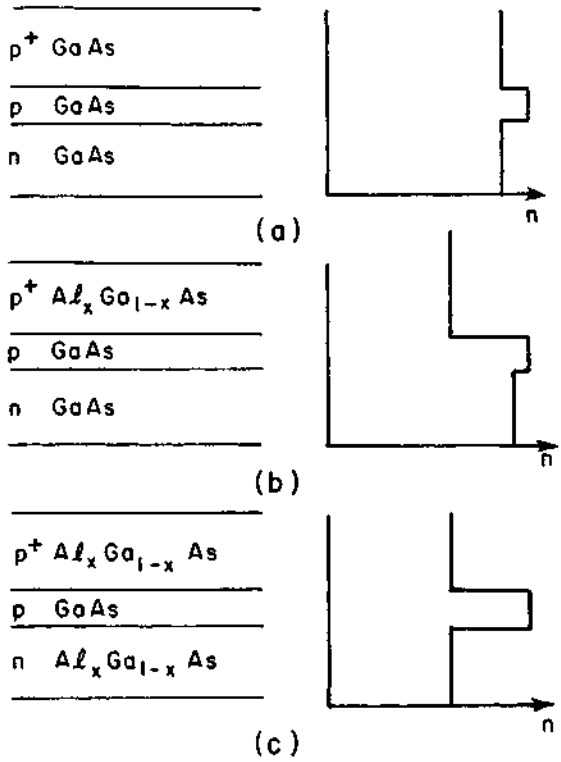


Figure 3-28 Structure and profile of index of refraction n for various types of junction in the aluminum gallium arsenide system. (a) Homojunction. (b) Single heterojunction. (c) Double heterojunction.

shown in Figure 3-24. The heterostructure reduces the light that leaks into the p^+ region because of waveguiding effects. This in turn leads to lower losses, lower current requirements, reduced damage, and longer lifetimes for the diodes. Single heterostructure diode lasers are still in use, but they have largely been replaced by double heterostructure devices.

Figure 3-28c shows the double heterojunction (or double heterostructure), so called because there are two changes in the composition of the material through the junction.. This confines the light on both sides by waveguiding effects. This reduces losses and lowers the current requirements even further. Double heterojunction devices are now widely used, having mostly replaced the other types. They offer the possibility of very long lifetime (many tens of thousands of hours) if used properly.

There are many different types of structure for semiconductor diode lasers. It is impossible to cover them all here. Figure 3-29 presents a simplified classification scheme showing some of the major subdivisions of diode lasers and their relationships.

The first major division is between edge emitting and surface emitting devices. So far, we have described only edge emitting structures. They are of the form shown in Figures 3-18, 3-26, and 3-27, in which the light emerges from the edge of the device, where the junction intersects the surface, in the plane of the junction. The configuration is simple and easy to fabricate. Most diode lasers are edge emitters. But edge emitters do suffer from the drawback that the volume of material that can contribute to the laser emission is limited, and they are difficult to package as two-dimensional arrays.

Recently, there has been considerable research activity aimed at development of surface emitting diodes, in which the light emerges from the surface of the chip rather than from the edge. This feature is attractive because devices could be packed densely on a semiconductor wafer and it is possible to fabricate two-dimensional arrays easily. In some devices, the laser cavity is vertical, that is, perpendicular to the plane of the junction. These devices are called vertical cavity surface emitting lasers (VCSELs). Some commercial models of surface emitting diode lasers have become available, and it seems certain that they will continue to develop in the future. Still, edge emitting lasers are far more common, and the remainder of this discussion will concern only edge emitting devices.

The edge emitting devices may be divided into homojunctions, single heterojunctions, and double heterojunctions, as we have discussed. The homojunctions are no longer used, and double heterojunctions dominate most applications.

Many semiconductor lasers have a stripe geometry, in which the gain region is confined to a narrow stripe region, as we have described. Without a stripe, the laser light will emerge all the way across the width of the diode; this is denoted in the figure as "Broad Area." It allows the generation of higher power, because a larger volume of material may contribute to the laser operation, but it leads to greater astigmatism, because the limiting apertures in the directions parallel and perpendicular to the junction have much different values.

There are two main approaches to defining the stripe geometry. Confining the laser operation within a stripe region may be accomplished by gain guiding or by

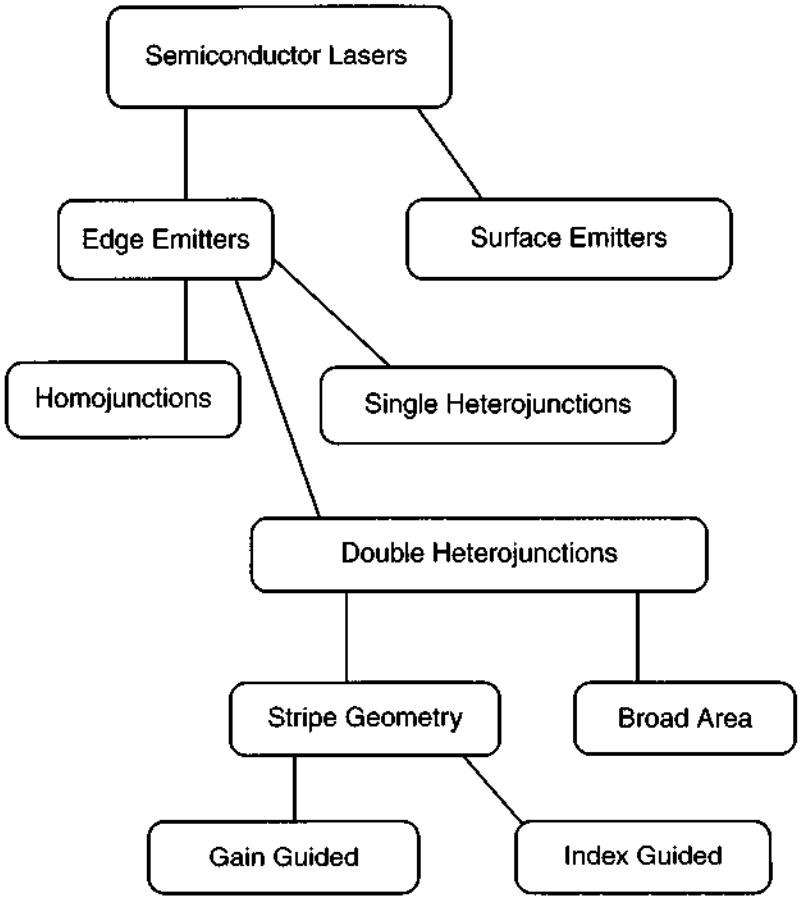


Figure 3-29 Simplified classification of major types of semiconductor laser structures.

index guiding. Both methods of confine the light so as to reduce the loss due to beam spreading in the junction plane. This in turn reduces the current requirements for laser operation and thus allows operation well away from the damage threshold. The use of some method of confining the light is essential to producing low-threshold, long-lived laser diodes. It also leads to limiting apertures that are more nearly equal in the directions parallel and perpendicular to the junction and thus reduces astigmatism.

Index-guided lasers employ steps in the index of refraction both parallel and perpendicular to the junction to confine the light. We have already discussed confinement perpendicular to the junction in the discussion of heterojunctions. Index-guided lasers also have a change in the index of refraction in the plane of the junction to reduce spreading of light within the junction. Figure 3-30 shows one geometry as an example, out of many possible variations.

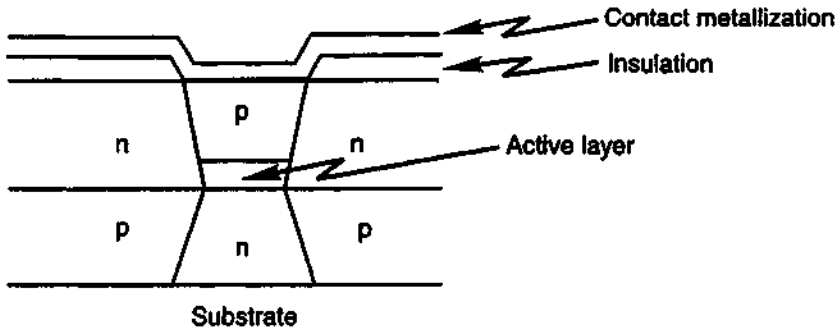


Figure 3-30 Example of an index-guided structure, called a buried heterostructure. The active region is GaAs. The notations *p* and *n* refer to regions of *p*- and *n*-type AlGaAs, respectively.

Gain guiding employs composition changes and refractive index changes in the direction perpendicular to the junction to confine the light, just as index-guided devices do. But the confinement of light in the plane of the junction is accomplished by refractive index changes that result from higher charge carrier density in the region where gain is high. The use of stripe geometry helps the gain guiding by confining the carriers and the gain region. Figures 3-26 and 3-27 have shown two examples of gain-guided structures, one in which the gain is confined by ion implantation to reduce the electrical conductivity in selected regions, and the other in which the current is confined by selective contacting of the metallization to the semiconductor.

Gain-guided lasers are easier to fabricate than index-guided devices, but they have weaker confinement, which leads to somewhat poorer beam quality and stability. Gain-guided lasers also have somewhat larger astigmatism than index-guided devices. They are also more difficult to operate in a single longitudinal mode, whereas often index-guided devices operate in a single longitudinal mode when the current is well above threshold. These factors may restrict the use of gain-guided devices for some applications where beam quality is important, but for many applications, gain-guided diode lasers are suitable.

To summarize the differences between gain-guided and index-guided structures, index-guided devices have small astigmatism, lower threshold current, and usually a relatively clean output beam, often a single transverse mode, but they are more complex and costly to fabricate. Gain-guided devices have larger astigmatism, higher threshold current, and usually multiple transverse modes in the output, but they have a simpler fabrication process and thus are less expensive. Both types are readily available commercially.

It is also worthwhile to consider some of the structures used to form the resonant cavity for a semiconductor laser. So far, we have assumed that the resonant cavity is formed by the polished or cleaved ends of the semiconductor material, in a configuration of two parallel mirrors, as indicated in Figure 3-18. This is the so-called

Fabry–Perot configuration. In fact, most resonant cavities for semiconductor lasers are formed in this way. But there are a number of other types of resonant cavity. The simple Fabry–Perot configuration usually leads to operation of the laser in several longitudinal modes, with resulting broad emission spectrum. For applications that require narrower spectral width, like lightwave communications, other types of cavities have been used. These types lead to single-mode output. These variations include the following:

Distributed Bragg reflector cavity. These devices use gratings as reflective elements. The grating structures are at the ends of the active region. They act as mirrors by scattering light back into the active medium. Such devices can operate in a single longitudinal mode with narrow linewidth.

Distributed feedback cavities. This cavity also uses a grating structure, but the grating is next to the active region and along the length of it. The grating scatters light back into the active region, so that the mirror is effectively distributed along the entire length of the laser cavity. This structure is more difficult to fabricate, because it requires growth of an epitaxial layer on top of the grating. Such lasers operate in a single mode with linewidth as narrow as 20 MHz. Distributed feedback semiconductor lasers are commercially available.

External cavity devices. These lasers use external optics to form the cavity. In one version, one reflector is a grating that may be rotated to tune the wavelength of emission, and the other reflector is a dielectric mirror. Such cavities can be much longer than conventional semiconductor laser cavities, which typically are of the order of 1 mm long. These devices can have very narrow spectral linewidth. They have some commercial availability.

Before concluding the discussion of laser diode structures, it is worth mentioning two additional types of structures that have developed rapidly in recent years. These are quantum well devices and high-power laser diode arrays. A quantum well is a very thin layer of semiconductor material between two layers with larger values of bandgap. An example is GaAs between AlGaAs layers. If the layer is thin enough, 20 nm or less, quantum mechanical properties of electrons become important. This changes the energy level structure of the material.

Quantum well devices may incorporate a single quantum well or multiple quantum wells, with a number of alternating thin layers of high- and low-bandgap material. The use of quantum wells in laser devices allows the material grower to optimize the properties of the material for the specific application. Quantum well devices offer lower threshold current and higher output power than devices without quantum wells.

Quantum well diode lasers have been the topic of intense research for a number of years. They are becoming commercially available.

The development of high-power linear diode laser arrays offers high-radiance diode sources suitable for applications like pumping solid state lasers. The array is fabricated as a bar with a number of stripe laser sources. A typical arrangement is

shown in Figure 3-31. The stripes are defined by proton implantation, which increases the electrical resistivity of the regions between the emitting areas. There is sufficient coupling of light between the stripes that the entire device acts as a single laser. These devices are capable of emission of many watts of optical power.

Such lasers are usually available as bar structures, typically 1 cm long. A single commercially available 1 cm bar can emit up to tens of watts of power in continuous operation, and research models have emitted hundreds of watts.

A number of bars may be stacked to form two-dimensional arrays. Figure 3-32 shows one approach to a stacked array. In this figure, the pitch between diode bars is shown as 0.4 mm. Other pitches are possible. These arrays are still expensive but represent an important, rapidly developing aspect of semiconductor laser technology. Commercial models are available with up to 50 linear arrays in the stack and are capable of emitting up to 5000 W in quasi-CW operation.

In summary, semiconductor laser structures can be extremely diverse. We have covered a few of the major variations in this discussion.

D. Organic Dye Lasers

Organic dye lasers represent a type of laser somewhat different from the more familiar gas and solid state lasers. They employ liquid solutions of complex dye materials. The dyes are large organic molecules, with molecular weights of several hundred. These materials are dissolved in organic solvents, like methyl alcohol. Thus, the active material is a liquid. Dye lasers are the only type of liquid laser that has reached a well-developed status.

One of the most important features that dye lasers offer is tunability. The monochromatic output of available dye lasers can be tuned over a broad range, from the ultraviolet to the near infrared. Dye lasers are chosen for applications in which tunability is important. They probably will not be used for applications in which one simply desires a visible laser output, but the exact wavelength is not important. For such applications (like alignment), other lasers, such as helium–neon, offer better convenience and economy.

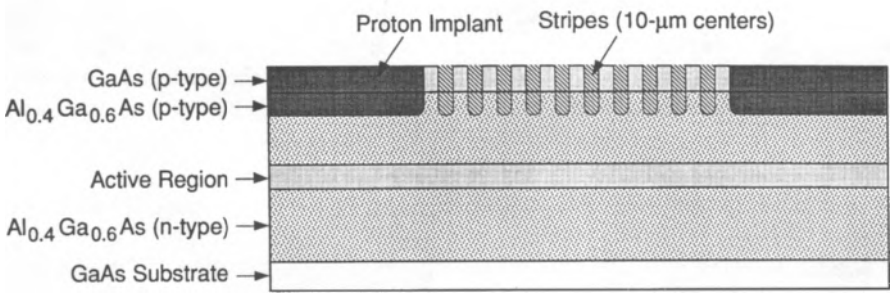


Figure 3-31 A multiple-stripe diode laser.

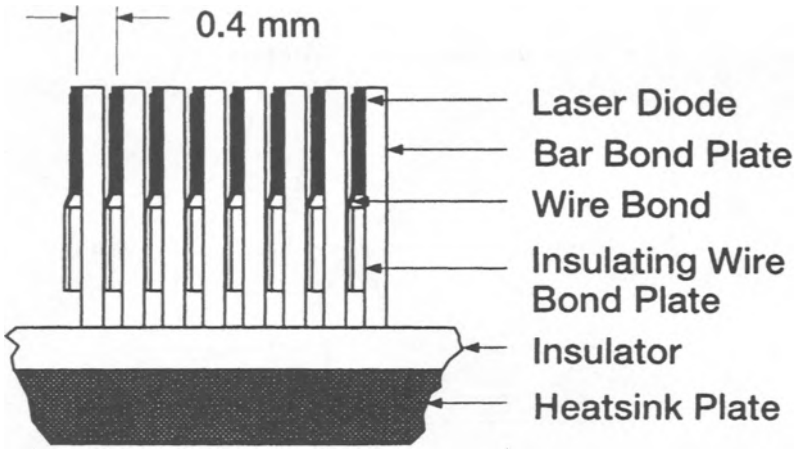


Figure 3-32 One approach to stacking of laser diode arrays. (Based on information from SDL, Inc.)

The dye materials that are used in lasers are similar to many familiar dyestuffs employed as colorants in fabrics, plastics, soaps, cosmetics, and so forth. Common dye materials often contain a chain of carbon atoms with alternating single and double bonds. The dyes are used in liquid solution at concentrations that are very low, often less than one gram per liter. It is characteristic of dye materials that small amounts are effective in producing significant changes in the optical properties of materials.

To introduce a specific dye material, let us consider the dye rhodamine 6G, one of the more important laser dye materials. It has several benzene rings and has the chemical formula $C_{26}H_{27}N_2O_3Cl$, with a molecular weight around 450. It is soluble in methyl and ethyl alcohols and in other organic solvents. It can be used to color silk or paper pink. Its absorption and emission spectra are shown in Figure 3-33. The emission and absorption are both broad, with the emission offset to somewhat longer wavelength. The broadness of the emission spectrum is what allows tuning of the output wavelength. Rhodamine 6G is typical of the materials that are used in dye lasers.

A typical energy level diagram for a dye material is shown in Figure 3-34. Initially, the entire population of molecules is concentrated at the bottom of the ground level, S_0. When the dye is irradiated with light of wavelength corresponding to the energy difference between S_0 and S_1, some of the ground state molecules are raised to S_1. These levels are shown as broad energy states containing many vibrational and rotational sublevels, so as to form a continuous band. Thus, the absorption and emission spectra are broad. When a dye molecule is raised to a position in S_1 by the pump light, it relaxes in a very short time to the lowest sublevels of S_1. This relaxation occurs by nonradiative transitions, which produce heat. The upper levels of S_0 are initially empty. Thus, one obtains a population inversion between the lower sublevels

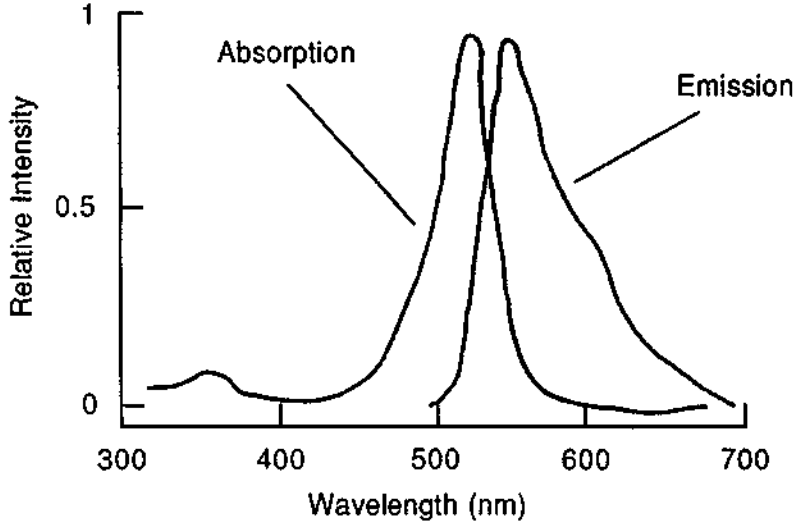


Figure 3-33 Absorption and emission spectra of rhodamine 6G.

of S_1 and the upper sublevels of S_0. These states are radiatively coupled; that is, the transition from S_1 to S_0 occurs with emission of fluorescent light. Because of the population inversion, gain by stimulated emission is possible, and laser action can occur. Because the laser transition occurs between the bottom of S_1 and a sublevel near the top of S_0, the laser light is at a wavelength longer than the pump light. Because many sublevels near the top of S_0 are empty, the laser light may be emitted over a range of wavelengths. This fact allows the possibility of tuning the output of dye lasers.

There is a competing process that reduces the number of dye molecules available for laser operation. Additional states, called triplet states and denoted in the figure as T_1 and T_2, are present. The states S_0 and S_1, which provide laser operation, are part of a different manifold of states, called singlet states. When dye molecules are excited to S_1, some of them make a transition to T_1 via collisions. The molecules that make this transition tend to be trapped because they remain in T_1 for a relatively long time. The decay from T_1 to S_0 proceeds slowly. As a result, T_1 acquires a large population. This reduces the number of molecules available for laser operation.

The situation is made worse by the presence of T_2. Absorptive transitions at the laser wavelength can take place between T_1 and T_2. It often happens that such triplet transitions occur at the same wavelength as the laser operation. Thus, as T_1 becomes populated, the medium becomes absorbing. This effectively shuts off the laser operation. The populating of T_1 occurs rapidly, within less than a few microseconds. This means that dye lasers require the use of very fast pumping sources to provide a short pulse of laser light before the triplet states become populated.

Thus, short-pulse lasers have often been used as the pump source. If the pump pulse is only a few hundred nanoseconds long, the triplet states do not become pop-

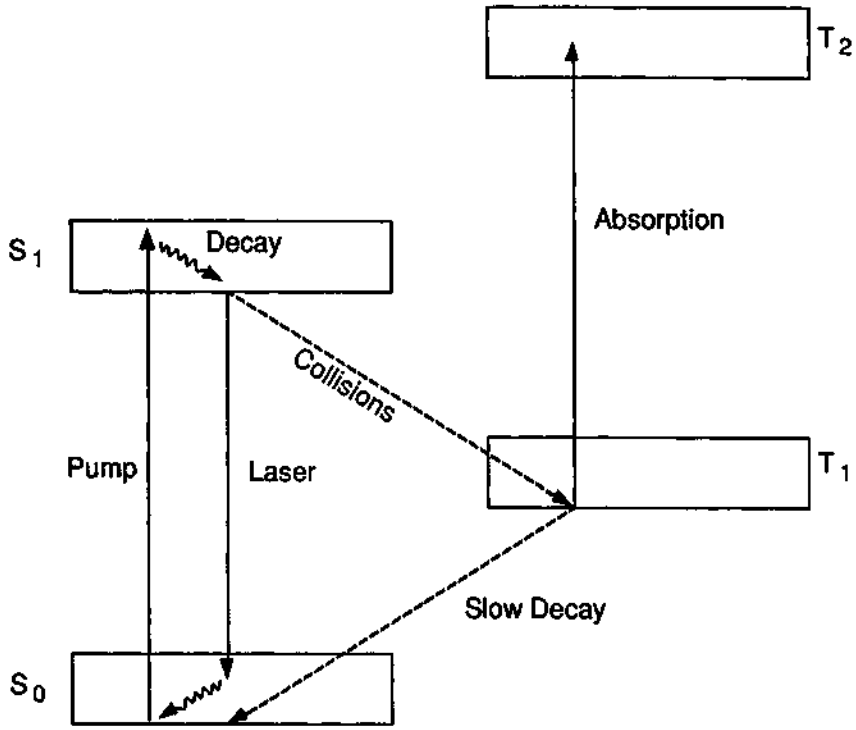


Figure 3-34 Energy levels relevant to operation of dye lasers. S_0 is the ground state, S_1 the first excited singlet state, and T_1 and T_2 the two lowest triplet states.

ulated, and the problem is eliminated. The short-pulse lasers that have been employed include nitrogen lasers operating at 337 nm, excimer lasers, and Q-switched, frequency-doubled Nd:YAG lasers. Then, the dye laser pulse often has duration in the nanosecond regime.

One can bypass the problem with the triplet states, and produce a continuous dye laser. The dye is rapidly circulated through a continuous pump beam. Each fluid element in the dye sees a brief pulse of pumping light as it passes through the pump beam, and it emits laser light. Then, as molecules begin to accumulate in T_1, the fluid element circulates out of the laser cavity and is replaced by fresh fluid. This in turn emits laser light briefly and then is replaced. The result is a dye laser with continuous output. The dye is circulated by a pump. The circulation time is long enough so that the molecules in T_1 will decay back to the ground state before the fluid element returns to the laser region.

Figure 3-35 shows a typical type of design. The pump source is an argon laser, the beam of which is focused to a small spot. The dye flows through the laser cavity. A typical flow velocity is 10 m/sec. If the pump beam is focused to a diameter around 10 μm, the duration of the effective pumping pulse as the dye molecules flow through the beam is around 1 μsec. The laser efficiency and output power

improve as the flow velocity increases, and as the diameter of the focal spot decreases. The dye flows in a high-velocity jet, which is tilted at Brewster's angle to the incoming pump light, so as to avoid losses by reflection at the surface of the stream. In the design shown in Figure 3-35, the curved pump mirror focuses the pump beam so that its beam waist coincides with the waist of the dye laser beam inside the dye stream.

In practice, most continuous dye lasers are pumped by argon lasers. The argon laser may operate at one of its strongest visible lines (488 or 514.5 nm), it may operate multiline in the visible, or it may operate in one or more ultraviolet lines. The dye laser output power and wavelength range vary according to which pump wavelength is chosen.

The design shown in Figure 3-35 uses a wavelength selecting birefringent filter as the tuning element. The birefringent filter consists of one to three crystalline plates of birefringent material, like crystalline quartz, inserted in the beam and tilted at Brewster's angle to the beam. The beam traveling through the plates is polarized and has two components of polarization, one parallel to the "fast" axis of the crystal and one along the "slow" axis. The two components travel at different velocities and become increasingly out of phase with each other. The first plate is cut to a thickness that results in a phase retardation of one full wavelength in the visible region. When white light travels through the first plate, only one wavelength is retarded by exactly one wavelength. We will refer to this as the preferred wavelength. This wavelength will emerge from the first plate with its polarization unchanged, but all other wavelengths will emerge with an elliptical polarization, that is, with a component of polarization perpendicular to the original direction. This perpendicular component is partially reflected from the Brewster's angle surfaces. This produces loss in the system for all wavelengths except for the preferred wavelength. These losses are sufficient to exceed the gain in the cavity, so that only the preferred wavelength can operate. Additional filter plates are often added to narrow the linewidth further. The second plate is twice the thickness of the first, and the third is four times the thickness of the first. A three-element filter is common.

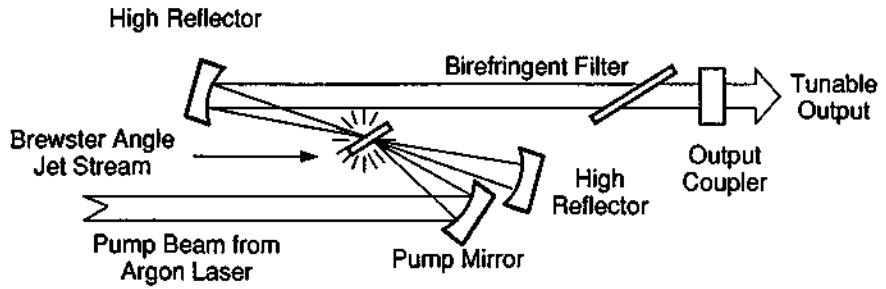


Figure 3-35 Tunable continuous dye laser using a flowing jet stream of dye. (Courtesy of Coherent, Inc.)

The birefringent filter is tuned by rotation about an axis perpendicular to the optical surfaces of the plates. The direction between the beam and the crystalline axes changes, so that the amount of retardation changes. The result is that the preferred wavelength changes for different angular orientations of the filter. For a given position of the filter, light of one particular wavelength will reach the mirror and be reflected back through the laser cavity. As the filter is rotated, the wavelength that can reach the mirror will change, and the wavelength of laser operation will change accordingly. Other types of tuning elements, such as prisms and diffraction gratings, have also been used.

Although not shown in Figure 3-35, many dye lasers incorporate one or two etalons in the form of parallel plates to select a single longitudinal mode.

The tuning range for any one particular dye that may be covered by rotating the tuning element is typically several tens of nanometers. The output power is highest near the middle of the range and drops off at either end. When the end of the tuning range is reached for a given dye, one may physically change the dye material. This is done by removing the reservoir containing the dye and replacing it with a reservoir containing a different dye. In this way, one may tune the laser over a broad range, covering the entire visible spectrum and parts of the infrared and ultraviolet. This capability for covering a broad wavelength range by the combination of tuning individual dyes and overlapping a number of dyes is illustrated in Figure 3-36, which presents results for a continous flowing dye laser pumped by an argon laser. At the short-wavelength end, one uses the ultraviolet lines of the argon laser. At longer wavelengths, the visible lines of the argon laser are used. Of course, if a different pump source were used, the form of the curves would shift somewhat.

As the figure indicates, one may tune between approximately 560 and 630 nm with rhodamine 6G as the dye material by rotating the tuning element. When the end of the tuning range of rhodamine 6G is reached, the rhodamine 6G is replaced with sulforhodamine B, and one may tune farther to the red. One changes the dye material each time the end of a tuning range is reached and obtains continuous coverage of the spectrum by overlapping the individual tuning curves.

The tuning range may be extended farther into the ultraviolet by frequency doubling the output of the basic dye laser. If one includes frequency-doubled models, commercial dye lasers are available to cover the spectral range from approximately 200 nm in the ultraviolet to 1036 nm in the near infrared. Some additional dyes are available that are capable of laser operation at wavelengths almost to 1200 nm.

The capabilities of commercial continuous flowing dye lasers are presented in Table 3-14. The method of pumping is indicated, and for each dye, the maximum value of output power near the middle of its tuning range is given. At a number of wavelengths, one can obtain output power of at least 6 W.

Pulsed dye lasers that are commercially available take a wide variety of forms. There are more models of pulsed dye lasers than continuous dye lasers, with many variations and many options. There are models pumped by excimer lasers, by frequency-doubled Nd:YAG lasers, by nitrogen lasers, and by flashlamps. The performance depends on the type of pump source used and the specific dye material

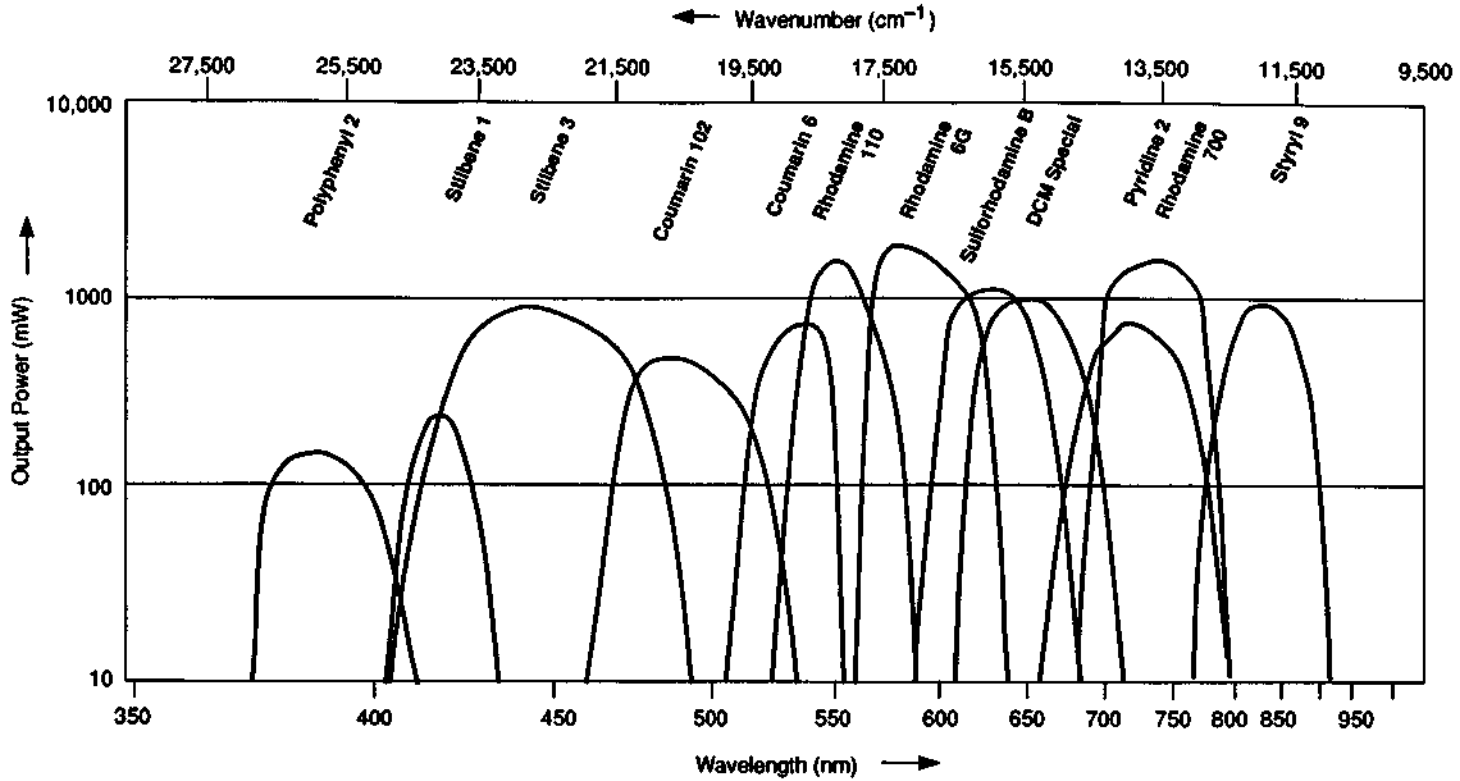


Figure 3-36 Tuning capabilities of a dye laser pumped by an argon laser. The name of the particular dye is given for each curve. (Based in information from Lambda Physik, Inc.)

Table 3-14 **Continuous Laser-Pumped Dye Lasers**[a]

Dye	Peak wavelength (nm)	Tuning range (nm)	Pump	Maximum power (W)
Polyphenyl 2	384	370–406	Ar^+, 300–336 nm	2.0
Stilbene 1	415	403–428	Ar^+, all UV lines	3.0
Stilbene 3	435	410–485	Ar^+, all UV lines	5.0
Coumarin 102	482	463–515	Ar^+, 350–386 nm	3.0
Coumarin 6	535	510–550	Ar^+, 488 nm	6.0
Rhodamine 110	550	530–580	Ar^+, 514.5 nm	6.0
Rhodamine 6G	575	560–625	Ar^+, 514.5 nm	6.0
Sulforhodamine B	625	598–650	Ar^+, 514.5 nm	6.0
DCM Special	645	610–695	Ar^+, 514.5 nm	6.0
Pyridine 2	720	675–783	Ar^+, 514.5 nm	7.5
Rhodamine 700	740	690–785	Kr^+, all red lines	4.6
Styryl 9	830	785–900	Ar^+, 514.5 nm	6.0

[a]Based on information from Lambda Physik, Inc.

employed. Table 3-15 presents some characteristics for pulsed dye lasers, in the wavelength region where the outputs are largest. What is presented are the optimum characteristics available in commercial models. For example, for some combinations of pump source, dye, and wavelength, it is possible to convert up to 30 percent of the pump energy into dye laser output, but a more typical value is perhaps around 10 percent.

Another variation of dye laser worth mentioning is the ring dye laser, illustrated in Figure 3-37. The cavity of the ring dye laser is formed by four mirrors. The light travels around the cavity in a ring. After it makes a complete circuit, the light then retraces its path. The optical isolator, to be described in Chapter 5, is a unidirectional device that allows light to travel in only one direction, as indicated by the arrow. The tuning element and dye jet are similar to what have been described earlier.

Table 3-15 **Pulsed Dye Laser Capabilities**

Parameter	Best values over 400–450 nm wavelength range
Energy per pulse	150 mJ
Peak power	20 MW
Average power	15 W
Efficiency of conversion of pump laser energy	30 percent
Pulse duration	2 nsec less than pump
Pulse repetition rate	1000 pulses per sec
Beam divergence angle	0.5 mrad
Spectral linewidth	1.2 GHz

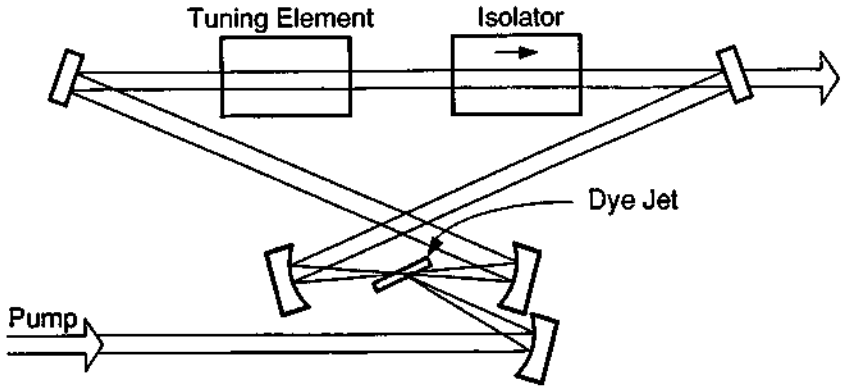


Figure 3-37 Ring dye laser.

The reason for using a ring dye laser is higher efficiency and the ability to produce higher output at a single frequency, as compared with a standard dye laser. In an ordinary laser cavity, the light bounces back and forth between two mirrors, producing a standing wave with nodes along the optical axis. The nodes are positions where the electric field of the laser is near zero. There is no stimulated emission at the positions of the nodes. The inability to utilize the excited dye molecules at the nodes lowers the efficiency of the system and wastes energy stored in the population inversion. In the ring system, the wave is a traveling wave. The traveling wave has no nodes and sweeps through the entire volume of dye uniformly, extracting the stored energy by stimulated emission. This both increases the efficiency and allows the buildup of the laser output as a single frequency. It is estimated that the single-frequency output of a ring dye laser may be up to 15 times that of a conventional cavity. Whereas the spectral width of a conventional dye laser may be of the order of 1 to 2 GHz, the spectral width of a ring dye laser is often in the 10–20 MHz range. In fact, for one actively stabilized ring dye laser system, a spectral linewidth of 0.5 MHz is claimed for short time periods.

This narrowing of the spectral width of ring dye lasers is another desirable feature. It allows spectroscopic measurements with higher resolution than is possible for lasers with broader linewidth.

Mode-locked dye lasers are capable of emitting extremely short pulses, in the femtosecond regime. The minimum pulse duration for a laser is equal to the reciprocal of the spectral width of the laser emission. Because the energy levels of the dye laser are very broad, the minimum pulse duration becomes very short. Commercial models of ultrashort dye lasers are available with pulse duration around 100 fsec, and experimental models have generated pulses only a few femtoseconds long. This is a remarkable result, because the pulse duration contains only a few cycles of the oscillating electromagnetic field. This type of dye laser has been very useful in scientific and diagnostic applications and for studies of ultrafast phenomena in chemical and biological kinetics.

One disadvantage of dye lasers is the degradation of the dye with time, partly because of photochemical effects caused by the intense pumping light. The dye must be removed and replaced with fresh dye solution. The time for degradation of the dye varies with many factors, including wavelength of the pump, the intensity of the pump, and the specific dye. In some circumstances, a dye solution may degrade within a few hours, whereas in other situations, a dye may not need to be replaced for many months.

The main advantage of dye lasers is their tunability over a broad spectral range. If one simply desired a laser and did not care about the exact value of its wavelength, one would not choose a dye laser. Dye lasers generally utilize another laser as a pump and thus tend to become large expensive systems.

The tunability of dye lasers has led to their widespread use in spectroscopy, an application in which they have excelled. The short-pulse models have been widely used in research in ultrafast phenomena. But even in spectroscopic applications, especially in the near infrared and the red end of the visible spectrum, dye lasers are being replaced by all-solid-state systems, like Ti:sapphire, which offer higher efficiency, smaller size, and better economics.

References

[1] C. Bibeau, *et al.*, *Appl. Opt.* **31**, 5799 (1992).
[2] N. P. Barnes, *et al.*, *IEEE J. Quantum Electron.* **QE-23**, 1434 (1987).

Selected Additional References

General Reference

J. Hecht, *The Laser Guidebook*, 2nd ed., McGraw-Hill, New York, 1992.

A. Gas Lasers

A. L. Bloom, *Gas Lasers*, Wiley, New York, 1968.
W. W. Duley, *CO_2 Lasers, Effects and Applications*, Academic Press, New York, 1976.
C. K. Rhodes, ed., *Excimer Lasers*, Springer-Verlag, Berlin and New York, 1984.
W. J. Witteman, *The CO_2 Laser*, Springer-Verlag, New York, 1987.

B. Solid State Lasers

W. Koechner, *Solid-State Laser Engineering*, 3rd ed., Springer-Verlag, Berlin and New York, 1992.

C. Semiconductor Lasers

D. Botez and D. R. Scifres, *Diode Laser Arrays*, Cambridge University Press, New York, 1994.
G. A. Evans and J. M. Hammer, *Surface Emitting Semiconductor Lasers and Arrays*, Academic Press, Orlando, FL, 1993.

H. Kressel and J. K. Butler, *Semiconductor Lasers* and *Heterojunction LEDs*, Academic Press, New York, 1977

Y. Suematsu and A. R. Adams, eds., *Handbook of Semiconductor Lasers and Photonic Integrated Circuits*, Chapman and Hall, New York, 1994.

D. Organic Dye Lasers

U. Brackmann, *Laser-Grade Dyes*, Lambda-Physik, Göttingen, Germany, 1986.

F. J. Duarte and L. W. Hillman, eds., *Dye Laser Principles*, Academic Press, Boston, 1990.

Kodak Laser Dyes, Eastman Kodak Co., Rochester, NY, 1987.

F. P. Schaefer, ed., *Dye Lasers*, Springer-Verlag, Berlin, 1990.

Chapter 4 | Trends in Laser Development

Chapter 3 discussed details of many of the practical types of lasers that are now available. The emphasis was on commercially available devices, rather than on experimental models. The chapter presented tabulations of the characteristics of commercial devices. In this chapter, we discuss prospects for future developments.

We begin by discussing conventional lasers, that is, types that are well developed and widely available. These include gas lasers such as helium–neon, argon, CO_2, solid state lasers like Nd:YAG, and semiconductor lasers. These lasers have been available for many years and have reached some level of maturity. For some of these types, the most important developments in recent years have involved reliability and improved lifetime, with modest increases in output power. Early in the history of lasers, a laser was a fragile device, requiring considerable care and maintenance. That situation has changed. Lasers have become durable, reliable, and reasonably economical to operate. One outstanding example is the semiconductor laser. Early models were very short-lived (hours). Advances in structures and fabrication have made them very long-lived (years) if they are treated properly.

Some types of lasers have not advanced substantially in recent years, and prospects for rapid future change do not appear likely. These include some of the gas lasers like helium–neon and argon, and some lamp-pumped solid state lasers like Nd:glass and ruby. These relatively mature lasers appear to have reached a plateau in their development. In some cases, they are being replaced by other laser types for established applications.

We note that there have been advances in some lamp-pumped lasers, like the vibronic solid state lasers discussed in Chapter 3. These advances seem likely to continue.

Two types of established lasers are rapidly advancing, with considerable research and development efforts devoted to them. These are semiconductor lasers and diode-pumped solid state lasers. We may expect substantial advances in these two classes, which we will describe in some detail.

We also describe classes of lasers that have been developed for some time, but that have not reached a status of wide commercial availability. These are chemical

lasers, free electron lasers, and x-ray lasers. We will describe a tunable device, the optical parametric oscillator, which is becoming commercially available. We will conclude the chapter with a discussion of the availability of tunable laser technology.

A. Semiconductor Lasers

Semiconductor lasers are the subject of extensive research and development worldwide. It is not possible to summarize all research efforts in semiconductor laser technology here; we do note that the worldwide efforts should improve the performance parameters of commercially available devices in the near future. There are many experimental efforts that have the goals of increasing available power, especially in the visible spectrum, of extending device lifetime, of improving beam quality, of providing blue and ultraviolet semiconductor lasers, and of improving characteristics for semiconductor lasers operating in the red portion of the visible spectrum.

There is extensive development in the area of vertical cavity surface emitting lasers (VCSELs), which were described briefly in Chapter 3. VCSELs offer the advantage of easy packaging in two-dimensional arrays on a chip. Many different approaches to fabrication of VCSELs are being investigated. VCSELs are beginning to be commercially available. It appears that more types of VCSELs should become available, and that use of VCSELs in selected applications should become common.

Other research areas with high interest include development of advanced quantum well structures, development of semiconductor lasers with a master oscillator, power amplifier structure in order to increase output power without damage, and continued development of multielement array structures, especially structures in which the light from the individual elements is coupled to form beams with narrow spectral linewidth and high beam quality.

Quantum well structures, in particular, mentioned in Chapter 3, represent an important advance. In a quantum well laser, one or more very thin layers of a semiconductor are sandwiched between layers of a semiconductor with wider bandgap. The layer thickness is of the order of 10 nm or less. The structure can be a single quantum well, with one layer of the narrow-bandgap material, or a multiple quantum well, with many layers. A multiple quantum well structure is sometimes referred to as a superlattice. Quantum well devices using materials in the AlGaAs system are becoming common. As the dimension of the layer approaches the wavelength of the charge carriers in the material, the properties of the material change. The properties of quantum well materials are different from those of bulk semiconductors of the same composition. The properties are affected by the confinement of carriers in the potential well defined by the larger-bandgap layers. Confinement of charge carriers in the quantum well increases the gain coefficient and reduces the threshold current for laser devices. The width of the quantum well provides tuning for the laser, because the effective bandgap of the material changes. Lasers based on quantum well structures have become commercially available, and their use should continue to increase in the future.

There is extensive effort devoted to development of semiconductor lasers operating in the blue and ultraviolet portions of the spectrum. In the blue portion of the spectrum, low-power laser sources have been limited to helium–cadmium lasers and air-cooled argon lasers. Both these sources are inefficient and costly, consume relatively large amounts of power, and are larger than desirable. The availability of an efficient compact semiconductor laser source in this region would open up new applications.

In particular, blue semiconductor lasers would be of great interest for optical data recording. The areal density of information that can be stored on an optical disk is inversely proportional to the square of the wavelength used. Most optical storage devices have used aluminum gallium arsenide lasers operating at wavelengths near 780 nm. If one had a blue semiconductor laser available, the density of information stored could be increased nearly fourfold. Considerations such as these are driving the development of short-wavelength semiconductor lasers.

One approach has been direct frequency doubling of diode laser output. In one system that has become commercially available, an 860 nm, continuous, single-mode laser diode is doubled to produce a 430 nm output. The doubling is performed by a nonlinear optical crystal in an external resonant cavity. The frequency of the input signal is locked to a mode of the cavity. The output is a TEM_{00} beam with power greater than 10 mW.

Another approach has been to use diodes of II–VI compounds to generate short-wavelength laser light directly. Devices based on materials such as zinc sulfide selenide or cadmium zinc selenide have been fabricated as double heterostructures and operated as lasers. Continuous room temperature operation at a wavelength around 490–510 nm has been achieved in laboratory devices, at power levels up to 10 mW. These blue-green semiconductor lasers are not yet commercially available in the mid 1990s, but they will likely become available soon.

Another recent development is the operation of laser diodes based on gallium nitride (GaN) [1]. The first GaN lasers operated at 417 nm, in the violet, and additional wavelengths of operation at 402 and 376 nm have been announced. This extends the range of operation into the ultraviolet. GaN offers excellent material properties for reliable, long-lived devices. It will undoubtedly become an important short-wavelength semiconductor laser material.

At the red end of the visible spectrum, $Al_xGa_yIn_{1-x-y}P$ semiconductor lasers continue to advance. The power levels available continue to increase, and the shorter-wavelength models keep appearing. These lasers are replacing helium–neon lasers in some applications.

The price of semiconductor diode laser arrays should continue to decrease. This will allow their use in a wider variety of applications.

B. Diode-Pumped Solid State Lasers

Lamp-pumped solid state lasers were developed early in the history of lasers and after some time reached a level of maturity. For a number of years, the technology

of solid state lasers did not advance much. But in recent years, solid state lasers have seen rapid development and change. This is mostly due to the development of semiconductor diode-pumped solid state lasers. Diode-pumped devices offer the capability for much higher efficiency and smaller size than lamp-pumped lasers of similar output. The efficiency in turn reduces the amount of waste heat that must be removed, so that the accessory equipment, like the chillers, also are much more compact. Commercial models of diode-pumped solid state lasers have been available for a number of years, and their output power has been increasing steadily.

Another important area involves advances in solid state laser technology at shorter wavelengths, especially in the green portion of the spectrum. For many years, the argon laser has been the dominant laser source in the green. But argon lasers are inefficient, large, and expensive. Advances in frequency-doubled diode-pumped Nd:YAG lasers provide a compact, efficient, and more economical source of green laser light. The availability of green diode-pumped solid state lasers should allow development of applications in areas such as display, medicine, and remote sensing that are difficult now.

We may expect to see diode-pumped frequency-doubled Nd:YAG lasers replacing argon lasers in areas where electrical power, size, and cooling are important issues. It is also likely that green diode-pumped solid state lasers will replace argon lasers for pumping Ti:sapphire lasers and will replace lamp-pumped Nd:YAG lasers for some micromachining applications. The extent to which green solid state lasers will replace other lasers in established applications will depend on how rapidly the price of diode lasers decreases.

The applications of green Nd:YAG lasers have been limited in some cases by what has been called "the green problem." The output of continuous frequency-doubled solid state lasers is subject to random fluctuations in output power, as we now describe.

Because the efficiency of frequency doubling increases with the optical power, doubling is usually performed with the doubling crystal in a resonant optical cavity to increase the power level. Frequency doubling may be performed with the doubling crystal within the original laser cavity (intracavity doubling) or with the crystal in an external resonant cavity, outside the cavity of the fundamental laser. In the intracavity case, the presence of high-intensity light confined within the laser cavity leads to efficient doubling. But often, more than one longitudinal mode will be present. The green problem arises when several longitudinal modes couple via the nonlinear interactions. As a result of these interactions, a strong amplitude modulation is imposed on the green output.

Research workers have devoted substantial efforts to the solution of this problem. The green problem may be solved in intracavity doubling if the laser is constrained to operate in a single longitudinal mode. The use of an external resonant cavity allows enhancement of the optical field at the position of the doubling crystal, leading to efficient doubling. It also allows separate optimization of the original laser cavity and the cavity in which doubling is done. Models of stable green Nd:YAG lasers have been developed using both intracavity doubling and external cavity doubling.

Although progress has been made, highly stable continuous green Nd:YAG lasers are still not available at output powers of more than a few hundred milliwatts, although laboratory devices emitting more than 3 W have been demonstrated. But the conditions for stable operation are sensitive. Small fluctuations can destabilize the output. Extremely precise thermal control is required to stabilize such lasers. Thus, advances in the development of stable high-power continuous green solid state lasers are likely to be slow.

The situation is different with pulsed green lasers. The Q-switched operation of green solid state lasers becomes simpler. Q-switched green solid state diode-pumped lasers are available with average power outputs at the multiwatt level. Potential applications for these devices include laser radar, micromachining, and environmental monitoring.

In addition, diode-pumped solid state lasers are moving into the ultraviolet. Laboratory demonstrations have yielded more than 1 W of continuous output from a frequency-quadrupled Nd:YAG laser at 266 nm. Although such lasers are not on the market yet, it seems likely that they will be in the near future. Such devices could have applications in microelectronic fabrication, optical disks, and medicine. Q-switched frequency-quadrupled Nd:YAG lasers operating at 266 nm are available with average power near 1 W.

Another research trend involves the development of microchip laser devices. Diode-pumped microchip lasers are very small, efficient, and easy to fabricate. An example of a possible strucrture for a diode-pumped Q-switched Nd:YAG microchip laser is shown in Figure 4-1. The Nd:YAG is in the form of a small slab with millimeter or submillimeter dimensions. It is pumped by a diode laser. The lithium tantalate Q-switch is also in the form of a small chip, located in close proximity to the Nd:YAG. The left mirror is high reflectance at 1060 nm and low reflectance at 810 nm. The center and right mirrors are partially transmitting at 1060 nm. This configuration uses flat mirrors and is very simple to fabricate. In practice, some microchip lasers have used one curved mirror to stabilize the cavity. The design

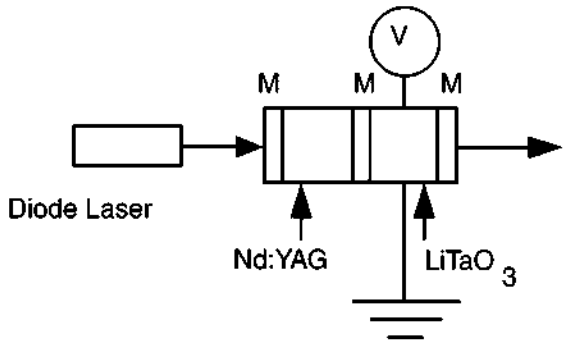


Figure 4-1 A diode-pumped, Q-switched Nd:YAG microchip laser. The mirrors are denoted M. The source of voltage for the Q-switch is denoted V.

shown in the figure could be considered typical, but there have been many variations, including frequency-doubled devices operating in the green.

Because the laser cavity is very short, there will be only one longitudinal mode within the gain curve of the material. This means that the microchip lasers are stable, single-mode devices.

The fabrication of microchip lasers is very simple, and well suited to mass production. Because of the easy fabrication and the small amount of material required, the cost can be low. Availability of microchip lasers as inexpensive, very small, stable, and efficient sources could enable new applications to become practical. Although some miniaturized diode-pumped solid state lasers are being marketed, the full potential of microchip lasers has not yet been realized. They should continue to develop.

Other developments in solid state lasers include new materials and advances in tunable devices. Tunable solid state lasers are replacing dye lasers for some spectroscopic applications.

In the area of new materials, solid state lasers based on materials with vibronic energy levels will continue to advance. Development of Ti:sapphire lasers will continue, especially in the area of very short (femtosecond) devices, useful for basic scientific studies. Lasers based on the materials Cr:LiCAF and Cr:LiSAF have been under development in a number of laboratories and should become commercially available soon. The long-wavelength end of the Cr:LiSAF absorption spectrum overlaps the emission of $Al_xGa_yIn_{1-x-y}P$ semiconductor diodes in the 670 nm region. Diode-pumped Cr:LiSAF lasers have been demonstrated and could form the basis of an all-solid-state laser system, although increases in efficiency and power are required.

Nd:YAG has been the dominant solid state laser material for many years. It probably will be challenged for use in some applications by the development of new solid state laser materials.

C. Chemical Lasers

Chemical lasers employ a chemical reaction to produce a population inversion. They offer the possibility of operation without an electrical input. All the required energy could be produced in the chemical reaction. One simply mixes chemical agents and allows them to react. In practice, most chemical lasers do use an electrical input in addition to the chemical energy released.

One of the leading examples of chemical lasers may be summarized by the following set of reactions:

$$\begin{aligned} F + H_2 &\rightarrow HF^* + H \\ H + F_2 &\rightarrow HF^* + F \\ HF^* &\rightarrow HF + h\nu \end{aligned} \tag{4.1}$$

In the first reaction, a free fluorine atom is required to initiate the reaction. Often, an electrical discharge is used to dissociate fluorine and produce some free fluorine atoms. The excited HF molecule, denoted HF^*, produced in the reaction is in an excited state, which is the upper level for the laser transition. In the second reaction, the free hydrogen atom from the first reaction interacts with fluorine to produce excited HF, leaving a free fluorine atom to continue in a chain of reactions. The third reaction indicates the transition of the HF molecule to its ground state, which is not populated in the chemical reactions. This is accompanied by emission of the energy difference as a photon. The population inversion is produced automatically as the chemicals react and yield some excited state molecules as their end product. Some electrical energy may be required for initiation, but once the reaction has begun, it can continue as long as the supply of reactants continues.

Table 4-1 shows some chemical laser systems that have been studied. Of these, the most highly developed are the hydrogen fluoride laser, operating at a variety of wavelengths around 3 μm, and the deuterium fluoride system, operating at a number of wavelengths in the 4 μm region. A few commercial models of these two types are available, with DF lasers capable of emitting up to 100 W in a TEM_{00} mode and HF systems capable of 60 W.

The chemical oxygen–iodine laser (COIL), operating at 1.32 μm, is of particular interest because it is potentially scalable to very high power. No commercial models are available, however.

Chemical lasers have been the subject of considerable research and development over many years. They can be scaled to very high powers, in excess of 100 KW in very large models. They have been of potential interest for military applications. Because their wavelengths are shorter than that of the CO_2 laser and their light is better absorbed by metals, they have been of interest for material processing applications. But the corrosive nature of the chemicals used in them has remained a problem, and they have never become used widely in industry.

D. Free Electron Lasers

Free electron lasers (FELs) represent a specialized class of device. They have received substantial publicity but probably have limited industrial applications. They

Table 4-1 **Chemical Laser Types**

Reactants	Laser molecule	Wavelength (μm)	Status
H_2–F_2	HF	2.6–3.6	Limited commercial availability
D_2–F_2	DF	3.6–5.0	Limited commercial availability
H_2–Cl_2	Hcl	3.5–4.1	Experimental
CS_2–O_2	CO	4.9–5.7	Experimental
D_2–F_2–CO_2	CO_2	10.6	Experimental
O_2–I_2	I	1.32	Experimental

use no solid, liquid, nor gaseous material as the active medium, Rather, the active medium is a beam of relativistic electrons. These electrons have velocity that approaches the velocity of light. A FEL directly converts the kinetic energy of the electron beam into light. The FEL is a novel type of laser with high tunability and potentially high power and efficiency.

The structure of a FEL includes a series of magnets called wigglers, which present an alternating magnetic field to the electron beam. Figure 4-2 shows a schematic diagram of a typical FEL structure. An electron beam is introduced into the laser cavity by magnets. The electrons radiate laser light by stimulated emission. The light emerges through the output mirror. The wavelength of the FEL is a function of a number of factors, including the velocity of the electron beam, the spacing of the wiggler magnets, and the magnetic field. The wiggler spacing may typically be a few centimeters, and the total device length a few meters. Once the device has been constructed, the wiggler spacing is fixed, but the laser wavelength may be tuned by varying the velocity of the electron beam. Operation of FELs at wavelengths from the ultraviolet to the millimeter wave regime has been demonstrated at a number of laboratories throughout the world.

Free electron lasers offer a number of potentially valuable properties, such as an essentially unlimited tuning range and the capability of being scaled to high power output. There is no material medium to be damaged. It is easy to extract heat from the laser, and the laser beam quality can be excellent. There has been high interest in FEL technology. Many laboratories and universities have constructed and operated FELs. Their results have received considerable attention throughout the laser community.

One important limitation for FELs is that they require a high-quality beam of relativistic electrons, with low angular spreading and very little variation in electron velocity. These sources are large and expensive. It requires substantial resources to build and operate an FEL. As an example, in the early 1990s, the so-called

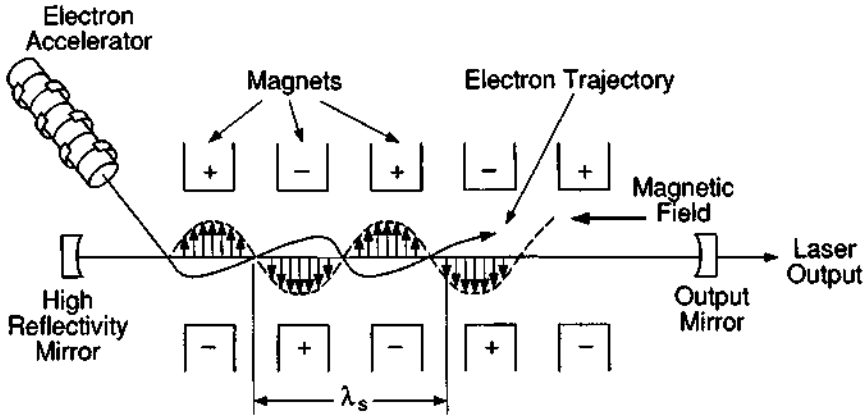


Figure 4-2 Free electron laser.

Advanced FEL, developed at Los Alamos National Laboratory, began operation after three years of design and construction [2]. In early operation, the laser operated in the 3–9 μm region, but it should be capable of operating in the visible and ultraviolet also.

This example indicates that FEL development is a major project. Thus, FEL technology will probably not come to widespread use in industrial applications. Potential uses for FELs include military applications, basic research at national laboratories, and possibly medical applications. Large electron beam capabilities exist in the nuclear medicine facilities at large medical complexes, and so these facilities could support the requirements of an FEL.

E. X-Ray Lasers

The desirability of lasers operating in the x-ray region of the spectrum has been known for many years. Many applications would be made possible by the availability of a coherent source of x-radiation. Relatively early in the history of lasers, there were some claims that x-ray laser sources had been developed. These claims are now believed to be spurious.

One problem is the lack of mirrors to form a resonant cavity for x-ray lasers. To date, x-ray lasers been superradiant devices, with only a single one-way passage of light through the material.

One active medium suggested for such lasers was the plasma produced by the interaction of high-power lasers with surfaces. Such plasmas are highly ionized and emit intense x-radiation. It was believed that a population inversion could be produced in some of the high-lying energy levels of multiply ionized species in the plasma.

In the mid-1980s, workers at Lawrence Livermore National Laboratory, using the very large Nd:glass lasers developed for laser-assisted thermonuclear fusion, detected stimulated emission in the 20 nm region when the laser was focused in a line on a thin selenium foil [3]. The emission came from 24 times ionized selenium (Se^{24+}) in the resulting plasma. The experiment involved amplified stimulated emission in a single pass through the plasma, because no mirrors were used. The most convincing evidence that this was laser emission was the fact that the intensity increased exponentially with the length of the plasma.

Since then, a number of different groups have demonstrated x-ray laser operation in highly ionized plasmas produced by very large lasers. The plasmas have involved multiply ionized elements such as germanium, tungsten, gold, yttrium, molybdenum, copper, and tin. The targets have most often been in the form of thin foils, which generate a linear plasma when heated explosively by the laser. The wavelengths have covered a range from 4 to 30 nm. Because the lasers were large and expensive, these demonstrations have not been compatible with practical applications.

In recent work [4], there has been a demonstration of x-ray laser operation in a C^{5+} plasma at 3.37 nm. The plasma was generated by a Nd:glass laser of tabletop

size. This result is significant because the size of the laser is more modest than in earlier experiments. It leads toward the possibility of x-ray lasers that do not require prohibitively large laser drivers. It is also important because the wavelength lies in the so-called "water window." This window (2.3–4.4 nm) lies between absorption edges of oxygen and carbon. It is potentially useful for biological studies. One could perform x-ray microscopy on living cells. This work could lead toward practical applications for x-ray lasers.

It is too early to predict what forms x-ray lasers will ultimately take and how extensive their applications may be. The earliest uses will be in the area of basic scientific research, rather than industrial applications. Applications that have been suggested include imaging of living cells, projection lithography with extremely small feature size, and x-ray holography.

F. Optical Parametric Oscillators

Optical parametric oscillators (OPOs) represent another tunable solid state source. These devices are based on nonlinear optical effects. We will discuss nonlinear optics more in Chapter 5, but to make this discussion self-contained, we will summarize some key points here.

In a noncentrosymmetric crystal illuminated by a laser beam, a large dielectric polarization, proportional to the square of the electric field of the laser, can be induced. The nonlinear polarization can radiate energy, permitting harmonic generation at multiples of the incident laser frequency. This is the basis of frequency doubling, tripling, and so on. If two laser beams are mixed in the crystal, radiation at the sum or difference in frequencies of the two beams may occur. In order for substantial buildup of the radiated energy at the new frequencies, a "phase matching" relationship must be satisfied. The wave vectors of the applied fields and the generated field must have the same relation as their frequencies. For sum-frequency generation, if the angular frequencies and wave vectors of the ith wave are ω_i and $\mathbf{k}_i$, respectively, and the subscripts 1, 2, and 3 refer to the first and second incident waves and the output, respectively, then

$$\omega_1 + \omega_2 = \omega_3 \qquad \text{and} \qquad \mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_3 \tag{4.2}$$

To satisfy both conditions simultaneously, one selects the crystal orientation, the direction of propagation, and the crystal temperature so that the birefringence of the crystal offsets the effects of dispersion.

The OPO uses nonlinear optical effects to generate two different frequencies, starting with one laser frequency. It is similar to difference-frequency generation. Strictly, an OPO is not a laser, but it generates a coherent optical beam like a laser, and it utilizes a resonant cavity like a laser. But there is no stimulated emission, so it is not a true laser.

In an OPO, a pump beam at frequency ω_p is passed through a suitable nonlinear crystal. Nonlinear interactions in the crystal transfer energy from the pump beam to a signal beam at frequency ω_s and at the same time generate an idler beam at fre-

quency $\omega_i = \omega_p - \omega_s$. The wavelength of operation may be varied by changing the temperature or the angular orientation of the crystal. A simplified configuration for an OPO is shown in Figure 4-3.

In this configuration, the output contains both the signal and idler. The left mirror is transmissive at the pump wavelength, and reflective at the signal and idler wavelengths. The right mirror is transmissive at the pump wavelength, and partially reflecting at the signal and idler wavelengths.

The device may be tuned by changing the phase matching conditions. The indices of refraction of the crystal at the three relevant wavelengths in the directions of beam propagation may be varied by changing the temperature of the crystal or by rotating the crystal. As a result, one obtains two beams with variable wavelength, the signal beam and the idler beam.

OPOs have been under development for many years and have become commercially available in the early 1990s. There has been substantial recent research interest in OPOs. Laboratory devices have demonstrated tuning from the ultraviolet to 14 μm in the far infrared. Other developments have included the first continuous devices and lower-cost technology.

Commercial devices tend to have short pulse durations (as short as femtoseconds) and are tunable in different ranges in the visible and near infrared. One device, for example, offers tuning of the signal from 410 to 690 nm and the idler from 730 to 2000 nm. If frequency doubling is used, the tuning range can be even wider.

OPOs represent a tunable all-solid-state device. Models available cover a broad wavelength range in the visible and infrared portions of the spectrum.

G. Tunable Lasers

In Chapter 3 and in the foregoing discussion, we have mentioned a number of tunable lasers. Because tunability is an important feature for many spectroscopic and photochemical applications, we summarize the availability of some leading tunable lasers in this section. It is beyond the scope of this chapter to enumerate them all.

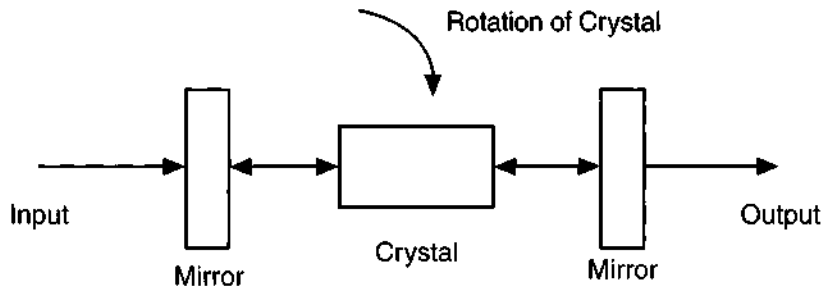


Figure 4-3 An optical parametric oscillator.

The use of nonlinear optical effects can increase the range of tunability for a tunable laser. For example, if one frequency doubles a laser tunable from frequency ω_1 to ω_2, the output is tunable from $2\omega_1$ to $2\omega_2$. If one mixes the output from a tunable laser in a nonlinear crystal with a fixed laser frequency, sum-frequency generation will yield a tunable output at a higher frequency and difference-frequency generation can yield an output at a lower frequency. Thus, use of nonlinear optical effects can greatly increase the spectral range of tunable lasers.

The leading tunable laser in the near ultraviolet, visible, and near infrared regions has been the dye laser, offering reasonably high power, narrow linewidth, and a broad tuning range. It has been employed for many studies of molecular structure and chemical reactions.

Tunable solid state lasers, such as Ti:sapphire, which offer a broad tuning range without the need to change dye materials, have begun to compete strongly in the visible and near infrared regions and have displaced dye lasers for some applications. With frequency doubling, they cover most of the range from 0.35 to 1.1 μm.

In the near infrared, $Al_xGa_{1-x}As$ and $In_xGa_{1-x}As_yP_{1-y}$ lasers and, in the long-wavelength infrared, lead compound semiconductor lasers are tunable by means of varying temperature and operating current. Many excellent spectroscopic studies have been performed with them. Any one device has a relatively limited tuning range. Thus, they are best suited for very high resolution spectroscopy within a narrow spectral range. In addition, mode hopping can make it difficult to reach any specified wavelength. Thus, it is desirable to use a semiconductor laser that does not exhibit mode hopping.

Raman shifting is a means that may be employed to extend the range over which a tunable laser may be tuned and to provide specific desired wavelengths that the laser could otherwise not reach. Raman scattering involves transfer of some of the energy of a photon to internal modes of a molecule. The frequency of the scattered light is shifted by a specific amount, characteristic of the molecular species. As a result, the range of available wavelengths may be expanded. Raman shifting has been used most often in experimental equipment, but some models are commercially available. The operation of Raman shifters will be described in the next chapter.

Optical parametric oscillators (OPOs) represent another tunable solid state source, based on nonlinear optical effects. They have been under development for a long time and are now becoming more widely available. They may be tuned by temperature or by rotating a crystal. Models available cover a broad wavelength range in the visible and infrared portions of the spectrum. Different commercial models cover at least the range from the blue portion of the visible spectrum to the mid infrared, near 4 μm. The short-wavelength end of the range can be extended by frequency doubling. For a single device, the tuning range can be relatively long. One commercial device, for example, covers the range from 1.45 to 4 μm. Most commercial models are short-pulse devices.

Finally, the free electron laser (FEL) offers potentially the ultimate in tunability, in principle being unlimited in its tuning range. Models of FEL devices have operated at wavelengths ranging from the ultraviolet to the millimeter wave regime. As

Table 4-2 **Tunable Lasers**

Type	Tuning range	Comments
Dye	UV to near IR	Needs another laser as pump; range includes frequency doubling
III–V semiconductor	Red and near IR	Tuning range small for any one device; tuning may be discontinuous
Lead salt semiconductor	Far IR	Expensive and cryogenic
Solid state vibronic laser	UV to near IR	Range includes frequency doubling
Optical parametric oscillators	Visible and IR	Needs pump laser
Nonlinear effects using variable-wavelength driver	UV, visible, IR	Needs tunable laser pump
Free electron laser	UV to millimeter wave	Large and very expensive

described earlier, a FEL represents a very large investment and probably will not be appropriate for routine spectroscopic applications.

Table 4-2 summarizes the most commonly used tunable lasers and their ranges of operation. In conclusion, a variety of tunable lasers are available with a broad variety of characteristics, and covering the wavelength range from the ultraviolet far into the infrared.

References

[1] S. Nakamura, *et al.*, *Japanese J. Appl. Phys.* **35**, L74 (1996).
[2] D. C. Nguyen, *et al.*, Paper CThN1, 1993 Conference on Lasers and Electro-Optics, Baltimore, MD, May 2–7, 1993.
[3] D. L. Matthews, *et al.*, *Phys. Rev. Lett.* **54**, 110 (1985).
[4] H. Aritome, et al., Paper CTuH6, 1993 Conference on Lasers and Electro-Optics, Baltimore, MD, May 2–7, 1993.

Selected Additional References

C. A. Brau, *Free-Electron Lasers*, Academic Press, San Diego, 1990.
F. J. Duarte, ed., *Tunable Laser Applications*, Marcel Dekker, New York, 1995.
J. Hecht, *The Laser Guidebook*, 2nd ed, McGraw-Hill, New York, 1992.

Chapter 5 | Laser Components and Accessories

In this chapter we shall describe several components and different types of accessory equipment that are commonly used either in laser fabrication or together with lasers as part of a system designed for a particular application. These devices represent separate fields of technology, some of which were well developed before the advent of lasers. In many cases, the demands of laser technology have spurred further developments in these associated fields. Because they are often closely connected with laser fabrication or with laser applications, a complete description of laser technology should include discussion of these related topics. These technologies are well developed in their own right, and so the description in this chapter must be somewhat abbreviated.

Specific components used in laser fabrication include mirrors, polarizers, Q-switches, and harmonic generators. Infrared transmitting materials are needed as lenses and windows for infrared lasers. Accessory equipment often used with lasers in a system designed for a specific application includes stable tables, focusing optics, detectors, modulators, beam deflectors, and scanners. To help describe these devices, we include a discussion of the fields of electrooptics, acoustooptics, and nonlinear optics.

A. Mirrors

In Chapter 1 we described the resonant cavity, which provides multiple reflections of light back and forth through the active medium. The resonant cavity is defined by two mirrors. We now describe what these mirrors actually are.

The mirrors that are most often used in laser fabrication are evaporated coatings consisting of multiple layers of dielectric materials, vacuum deposited on a transparent substrate. Early in the history of lasers, thin metallic coatings, such as silver or aluminum, were often used. They were prepared by either vacuum or chemical deposition and, in the case of solid state lasers, could be directly on the end of the

laser rod. However, these mirrors inevitably had large scattering loss and were easily subject to damage. Almost all mirrors now used in practical lasers are multilayer dielectric mirrors, which can be very resistant to damage.

These mirrors use the principle of interference to produce high values of reflectance. They consist of alternating layers of materials with high and low index of refraction, as shown in Figure 5-1. Each layer has an optical thickness equal to one-quarter of the wavelength at which the mirror is to be used. The mirrors are prepared by vacuum evaporation, coating the substrate alternately with layers of the two materials.

Usually, a mirror has an odd number of layers, with the high index material next to the substrate. The reflectivity R of such a set of k layers is

$$R = \frac{n_2^{k+1} - n_0 n n_1^{k-1}}{n_2^{k+1} - n_0 n n_1^{k+1}} \tag{5.1}$$

where n_2 and n_1 are respectively the indices of refraction of the high- and low-index layers, $n_0 \approx 1$ is the index of refraction of air, and n is the index of refraction of the substrate. Figure 5-2 shows a plot of the reflectivity as a function of the number of

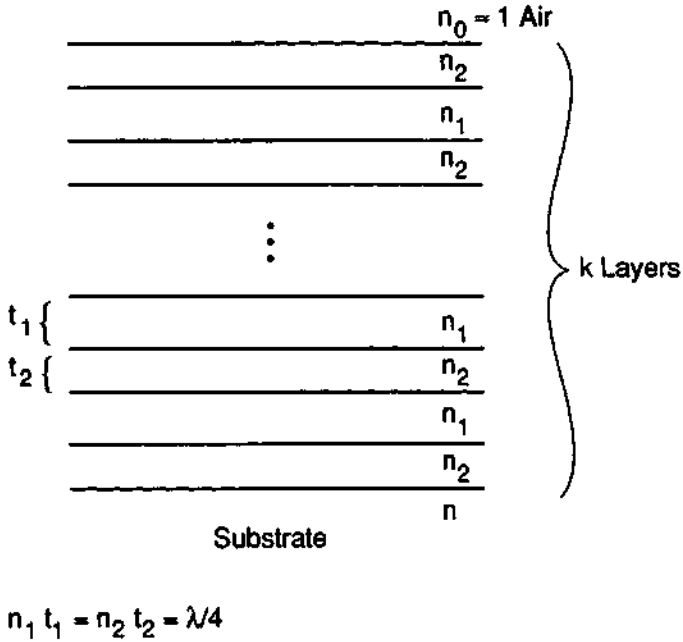


Figure 5-1 Schematic diagram of a multilayer coating, in which n_1 and n_2 are the indices of refraction of the two materials and t_1 and t_2 are their physical thicknesses. The optical thickness (equal to $n_i t_i$ for layer i) is equal to one-quarter wavelength. The index of refraction of the substrate is n, and the index of refraction n_0 of air is approximately unity.

layers in the stack for several different values of the index of refraction. The lower value is chosen as 1.46, representative of SiO_2. Various values for the high index are indicated. The value 2.62 is representative of titanium dioxide.

The figure illustrates that the reflectivity may be very high (~100 percent) for a reasonably small number of layers. Also, the reflectivity may be adjusted to any desired value by choosing materials with appropriate indices of refraction and by adjusting the number of layers. Scattering losses in carefully prepared mirrors may be less than 0.01 percent. Thus, mirrors with reflectivity exceeding 99.99 percent have been prepared, at least for He–Ne lasers. The mirrors for He–Ne lasers have been the subject of intensive development and probably are better developed than the mirrors for other lasers.

Multilayer dielectric mirrors are more durable than metallic mirrors and are the mirrors commonly used with most lasers. A wide variety of dielectric materials have been used for the coatings. The materials have included silicon dioxide, titanium dioxide, silicon monoxide, cerium oxide, zirconium dioxide, zinc sulfide, zinc selenide, cryolite (sodium aluminum fluoride), magnesium fluoride, and thorium fluoride. Silicon dioxide/titanium dioxide multilayer stacks are often used as laser mirrors in the visible spectrum.

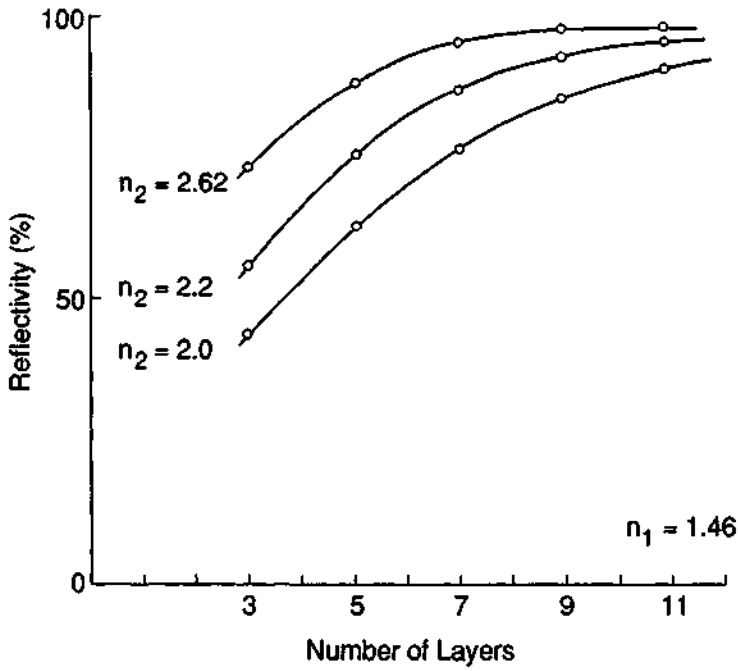


Figure 5-2 Reflectivity as a function of the number of layers in a multilayer dielectric stack for different values of the high index n_2 and with the low index n_1 equal to 1.46, representative of SiO_2.

The reflectivity as a function of wavelength for the mirrors for a typical small He–Ne laser is shown in Figure 5-3, for both the high-reflectivity mirror and for the output mirror, which has somewhat lower reflectivity. Because the He–Ne medium has relatively low gain, it cannot tolerate much loss, and thus the reflectivity of the output mirror is relatively high, over 98 percent in this example. Laser media with higher gain will use output mirrors with lower reflectivity.

There is one instance in which bulk metallic mirrors are commonly used. This is the high-reflectivity mirror in far infrared lasers, especially CO_2 lasers. Because the mirrors must withstand high irradiance, they may be relatively massive and are made of materials with high thermal conductivity. Internal channels for liquid cooling have sometimes been used. In high-power lasers, such cooling reduces thermal distortion and minimizes damage from the high power.

The metals most commonly used are high-purity copper and molybdenum. Copper mirrors can withstand higher irradiance without damage than molybdenum, whereas molybdenum suffers less thermal distortion than copper. The surface of high-purity copper is subject to tarnishing, and so the copper may be coated with a film of another metal, such as gold or electroless nickel.

All the mirrors we have described have undergone substantial development to meet the requirements of laser technology. They may be obtained from a number of suppliers, with a choice of radii of curvature for the different configurations of the resonant cavity described in Chapter 1.

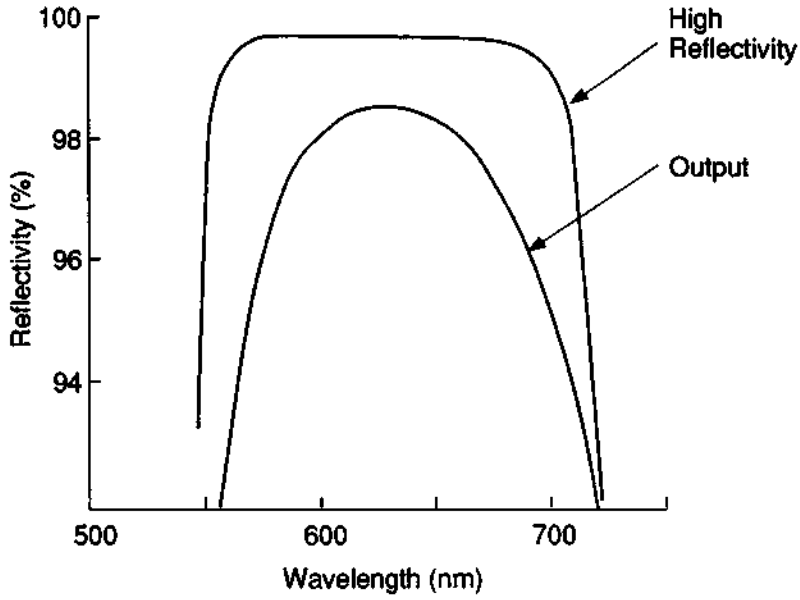


Figure 5-3 Reflectivity as a function of wavelength for mirrors for a typical He–Ne laser.

B. Optics

Many laser applications require optical elements to focus or shape the beam. Chapter 2 described the focusing properties of laser light, including the minimum diffraction-limited focal spot size and the effect of spherical aberration. The discussion in that chapter was limited to single-element lenses (singlets) with spherical surfaces. Here, we will describe some of the choices available for optical elements for use with lasers, and discuss how they may improve the focusing for laser applications.

As we saw, in many cases singlet lenses yield focal spots with diameter dominated by spherical aberration, rather than by diffraction. In such cases, the focal diameter may be reduced by use of multielement lenses or lenses with specially formed surfaces that are not spherical (aspheric lenses). The simplest case is that of a two-element spherical lens (doublet). A properly designed doublet lens can reduce the distortion of the wavefront as it passes through the lens, and it can also substantially reduce the effect of aberrations that increase the focal diameter above the diffraction limit. The cost of doublet lenses is higher than that of singlets, but the performance advantages may often be worth the added cost.

Calculations of the performance of multielement lenses become more complicated than for singlet lenses, and they usually are performed with computerized optical design software. An example of such a calculation is shown in Figure 5-4, which shows the fraction of the laser energy contained within a spot of specified radius, as a function of radius, for both a singlet and a doublet lens. It is apparent that the doublet produces a smaller focal diameter, containing the laser energy within a substantially smaller radius. The performance of the doublet is very near diffraction limited. Of course, use of multielement lenses cannot reduce the focal spot size below the diffraction limit.

Doublet lenses used with lasers should generally be air-spaced, rather than cemented lenses. The high laser irradiance can easily damage the cement.

Another option is the use of aspheric lenses, which can give excellent performance and which theoretically may be designed to eliminate spherical aberration completely. But aspherics are difficult to polish accurately, and their cost becomes very high. Aspheric optics may now be fabricated by a computer-controlled diamond turning process, and they are becoming more readily available than in the past. Still, aspheric lenses are usually employed only in specialized cases where a high degree of control of wavefront errors is needed.

Another advance in optics for laser systems involves diffractive optics. Diffractive optics employs the physical phenomenon of diffraction to manipulate optical beams, rather than the phenomena of refraction or reflection, which have been conventionally employed in optical systems.

Diffractive optics is a rapidly developing technology, which is now being incorporated in many optical designs, replacing refractive and reflective optical elements. In many cases, the use of diffractive optics offers reduced size and weight and improved performance compared with conventional optics. Use of diffractive optics also can reduce the number of components in an optical system, with attendant

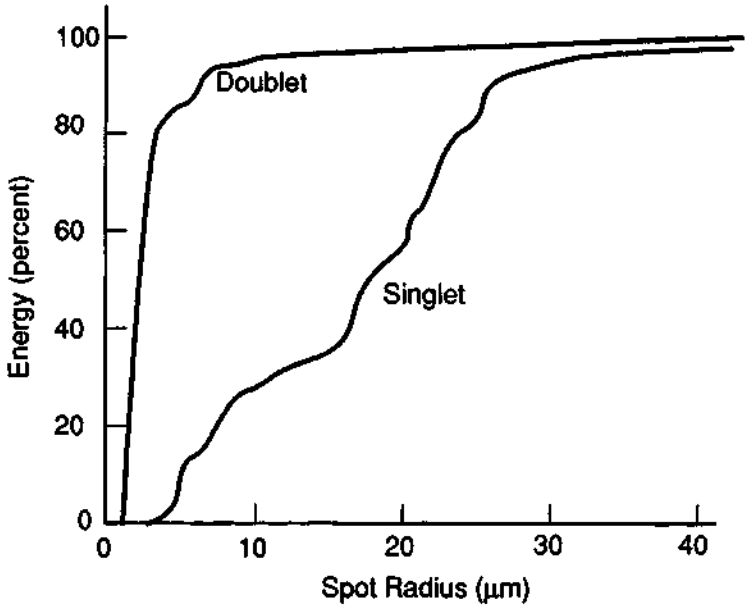


Figure 5-4 Comparison of the amount of laser energy contained within a focal spot of specified radius, as a function of radius, for a singlet lens and a doublet lens. The calculations are relevant to lenses with 2 cm clear aperture and 10 cm focal length. (From W. H. Lowrey and W. H. Swantner, *Laser Focus World,* p. 121 (May 1989).)

reduced cost. The applications of diffractive optics are rapidly increasing. They include use in head-mounted displays, and coupling light from laser diodes into optical fibers.

Diffractive optic lenses are being used for focusing laser beams onto the surfaces of workpieces. They seem to be used most often with CO_2 lasers in applications like welding, cutting, and scribing. They can reduce the aberrations associated with use of spherical lenses and produce a smaller focal spot and higher irradiance at the workpiece. Many users claim increased productivity from a laser system when diffractive optical elements are employed.

C. Polarizers

Elements to control the polarization of light are often used along with lasers. The most widely known type of polarizer is the so-called dichroic polarizer, which consists of a polarizing material embedded in a plastic sheet. Dichroic polarizers exhibit high transmission for light polarized linearly in one direction and high absorption for light polarized perpendicular to that direction. When a polarizing sheet is placed in a beam of unpolarized light, the transmitted light is linearly polarized and has

intensity one-half that of the incident light. The dichroic polarizing sheets are inexpensive and are often used in classroom demonstrations of polarization. The dichroic sheets do have some limitations. Because they absorb half the incident power, they can be used only with low-power lasers. The quality of the polarized light may not be sufficient for exacting laser work. The extinction ratio, defined as the ratio of the transmitted power with the selected polarization to the transmitted power with the perpendicular polarization, is not so high as it is with high-quality crystal polarizers.

One may obtain higher extinction ratios and greater resistance to damage by employing crystalline polarizing prisms, which use birefringence. Crystalline polarizing prisms are of high optical quality and offer high extinction ratios (>1000) between the two components of polarization. The crystal polarizers are more expensive than sheet polarizers, however.

The phenomenon of birefringence has been described in Chapter 1. Crystal polarizers operate by physically separating the two different components of polarization. This is possible because of the difference in the velocities of the two components. Figure 5-5 shows one type of polarizing prism. The crystal is split into two

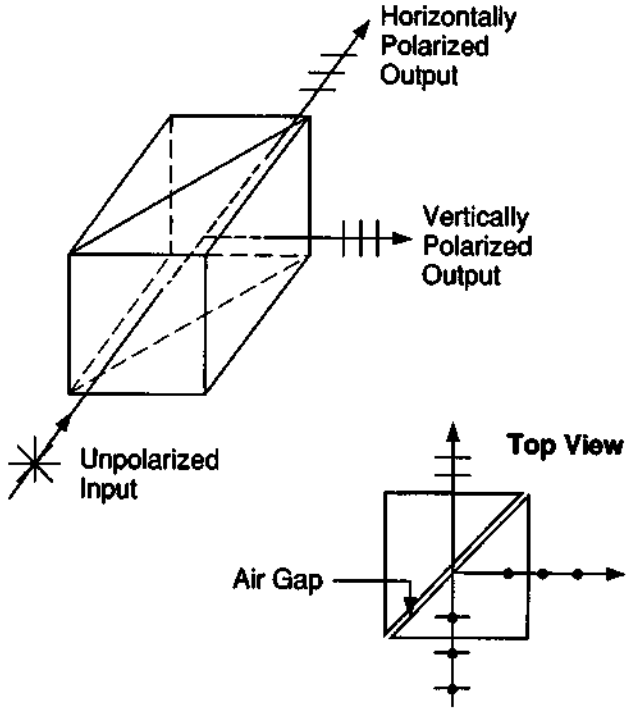


Figure 5-5 A polarizing prism in which a birefringent crystal is split into two triangular parts, separated by an air gap.

triangular parts, separated by a small air gap. Both components enter the crystal and travel in the same direction, but they have different velocities and different indices of refraction. When the light reaches the air gap, one component has index of refraction large enough so that the angle between the beam and the surface normal exceeds the angle for total internal reflection. This component is reflected, as indicated. The other component strikes the surface at an angle less than its critical angle and is transmitted, as indicated. The device is constructed with two triangular pieces of material so that the transmitted beam will not be deviated. The crystalline material, with its two different indices of refraction, and the angle of the interface are chosen so that one component is transmitted and the other is reflected as we have described. There are many other designs for polarizing prisms, but this discussion illustrates the general principle.

Some polarizers have optical cement between the two pieces. Generally, cemented prisms are unsuitable for work with lasers, because high irradiance damages the cement. Thus, air-spaced polarizers are usually used with lasers.

D. Infrared Materials

In many applications of lasers, the beam must pass through a transparent window or be focused by a lens. If the laser wavelength is in the visible, near ultraviolet, or near infrared portions of the spectrum, familiar materials may be used for the window or lens. Glass and quartz are transmissive in these regions. One may use the easy availability and well-developed technology of glass and quartz. Farther in the infrared, glass and quartz are no longer transmissive, and one must use other materials. The long-wavelength limit of transmission is near 2.7 μm for many optical glasses, and near 4.5 μm for quartz. Thus, glass or quartz may be used with Nd:glass and Nd:YAG lasers, but not with CO_2 lasers.

There are a number of optical materials that are transmissive farther in the infrared, particularly at 10.6 μm. Because the technology of such materials is less well known than that of glass and quartz, it is worth reviewing the availability of materials useful as windows and lenses at 10.6 μm. These materials include alkali halides and a number of semiconductors. Some commonly used materials are listed in Table 5-1. In the table, the long-wavelength limit refers to the longest infrared wavelength at which the material has useful transmission. The materials are ordered from top to bottom approximately in ascending order of both cost and performance.

The alkali halides, notably sodium chloride and potassium chloride, would appear at first to be ideal. They offer low absorption at 10.6 μm. They have low index of refraction, so that reflection losses are small. Their cost is low. Also, because they are transparent in the visible, an infrared system employing these optics can be lined up with visible light. The main disadvantage, which greatly restricts their use, is that they absorb water vapor from the atmosphere. This will cloud the surface, so that, unless special care is employed to provide a low-humidity environment, their life will be short.

Table 5-1 **Window and Lens Materials for CO_2 Lasers**

Material	Long-wavelength limit (μm)	Advantages	Disadvantages
NaCl and KCl	12, 15	Very low absorption, low cost, low index visible transmission	Poor stability, water absorption
Ge	23	Low absorption, good stability	High cost, high index, thermal runaway
GaAs	18	Low absorption, good stability	High index, very high cost
CdTe	30	Low absorption, good stability	High index, very high cost
ZnSe	20	Very low absorption, good stability, visible transmission	High index, very high cost

A number of semiconductor materials are useful, including Ge, GaAs, CdTe, and ZnSe. All of these materials are available with low absorption at 10.6 μm. All of them are opaque in the visible, except ZnSe. All these materials have high cost, with Ge currently being lowest. The cost increases rapidly as the size of the element increases. All four materials have high index of refraction and require antireflection coating to reduce losses. The semiconductor materials are subject to a phenomenon called thermal runaway, in which absorption increases with increasing temperature. Thermal runaway can lead to catastrophic failure if one tries to transmit too high power. Ge is particularly susceptible to thermal runaway; the other semiconductor materials are less affected. Thus, Ge optics may have to be water cooled. Of these four semiconductor materials, ZnSe probably offers the best combination of absorption coefficient and thermal conductivity for use as a transmitting element for high-power CO_2 laser beams. It is available with very low absorption in the form of chemically vapor-deposited material,

In summary, there are a variety of infrared transmitting materials suitable for use as optics for CO_2 lasers. One can make a choice based on a trade-off between cost and the performance requirements for a particular application.

E. Detectors

Detectors suitable for monitoring laser power are commonly employed along with lasers. For an application such as lightwave communications, a detector is necessary as the receiver. For applications involving interferometry, detectors are used to measure the position and motion of the fringes in the interference pattern. In applications involving material processing, a detector monitors the laser output to ensure reproducible conditions. In most applications, one desires a detector to determine

the output of the laser or the amplitude of a transmitted or received signal. Thus, good optical detectors for measuring laser power and energy are essential.

The detection and measurement of optical and infrared radiation is a well-established area of technology. In recent years, this technology has been applied specifically to laser applications, and detectors particularly suitable for use with lasers have been developed. Commercial developments have also kept pace. Detectors specially designed and packaged for use with lasers are marketed by numerous manufacturers. Some detectors are packaged in the format of a power or energy meter. These devices constitute a complete system for measuring the output of a specific class of lasers, and they include a detector, housing, amplification if necessary, and a readout device.

In this section, we will describe some of the detectors that are available for use with lasers. We shall not attempt to cover the entire field of light detection, which is very broad. Instead, we shall emphasize the detectors that have found most practical application with lasers. We shall also define some of the common terminology.

There are two broad classes of optical detectors: photon detectors and thermal detectors. Photon detectors rely on the action of quanta of light energy to interact with electrons in the detector material and to generate free electrons. To produce such effects, the quantum of light must have sufficient energy to free an electron. The wavelength response of photon detectors shows a long wavelength cutoff. When the wavelength is longer than the cutoff wavelength, the photon energy is too small to liberate an electron, and the response of the detector drops to zero.

Thermal detectors respond to the heat energy delivered by the light. The response of these detectors involves some temperature-dependent effect, like a change of electrical resistance. Because thermal detectors rely only on the amount of heat energy delivered, their response is independent of wavelength.

The performance of optical detectors is commonly described by a number of different figures of merit, which are widely used throughout the field of optical detection. They were developed originally to describe the capabilities of a detector in responding to a small signal in the presence of noise. As such, they are not always pertinent to the detection of laser light. Often in laser applications, for example, in laser metalworking, there is no question of detection of a small signal in a background of noise. The laser signal is far larger than any other source that may be present. But in other applications, like laser communications and detection of backscattered light in laser Doppler anemometry, the signals are small and noise considerations are important. It is also worthwhile to define these figures of merit because the manufacturers of detectors usually describe the performance of their detectors in these terms.

The first term that is commonly used is responsivity. This defines how much output one obtains from the detector per unit of input. The units of responsivity are either volts/watt or amperes/watt, depending on whether the output is a voltage or an electric current. This depends on the particular type of detector and how it is used. We note that the responsivity does not give any information about noise characteristics.

The responsivity is an important characteristic, which is usually specified by the manufacturer, at least as a nominal value. Knowledge of the responsivity allows the user to determine how much detector signal will be available in a specific laser application. One may also characterize the spectral responsivity, which is the responsivity as a function of wavelength.

A second figure of merit, one that depends on noise characteristics, is the noise equivalent power (NEP). This is defined as the radiant power that produces a signal voltage equal to the noise voltage of the detector. Because the noise depends on the bandwidth of the measurement, that bandwidth must be specified. The equation defining NEP is

$$\text{NEP} = \frac{HAV_N}{V_S(\Delta f)^{1/2}} \tag{5.2}$$

where H is the irradiance incident on the detector of area A, V_N is the root mean square noise voltage within the measurement bandwidth Δf, and V_S is the root mean square signal voltage. The NEP has units of watts per (hertz to the one-half power), commonly called watts per root hertz. From the definition, it is apparent that the lower the value of the NEP, the better the detector's ability to detect a small signal in the presence of noise.

The NEP of a detector depends on its area. To provide a figure of merit under standard conditions, a term called detectivity is defined. Detectivity is represented by the symbol D^*, pronounced D-star. It is defined by

$$D^* = \frac{A^{1/2}}{\text{NEP}} \tag{5.3}$$

Because many detectors have NEP proportional the the square root of their area, D^* is independent of the area of the detector and provides a measure of the intrinsic quality of the detector material itself, independent of the area with which the detector happens to be made. The dependence of D^* on the wavelength λ, the frequency f at which the measurement is made, and the bandwidth Δf are specified in the notation $D^*(\lambda, f, \Delta f)$. The reference bandwidth is generally taken as 1 Hz. The units of D^* are centimeters (square root hertz) per watt. A high value of D^* means that the detector is suitable for detecting weak signals in the presence of noise.

All optical detectors respond to the power in the optical beam, which is proportional to the square of the electric field. They are thus called "square-law detectors." Microwave detectors, in contrast, can measure the electric field intensity directly. But all the detectors that we consider here exhibit square-law response. This is also true of other common detectors such as the human eye and photographic film.

We shall now describe the operation of photon detectors. As mentioned before, photon detectors rely on liberating free electrons and require the photon to have sufficient energy to exceed some threshold. We will consider three types of photoeffect that are often used for detectors. These are the photovoltaic effect, the photoemissive effect, and the photoconductive effect.

1. PHOTOVOLTAIC EFFECT AND PHOTODIODES

The photovoltaic effect and the operation of photodiodes both rely on the presence of a *p–n* junction in a semiconductor. When such a junction is in the dark, an electric field is present internally in the junction region because there is a change in the level of the conduction and valence bands. This change leads to the familiar electrical rectification effect produced by such junctions.

When light falls on the junction, it is absorbed and, if the photon energy is large enough, it produces free hole–electron pairs. The electric field at the junction separates the pair and moves the electron into the *n*-type region and the hole into the *p*-type region. This leads to a change in voltage, which may be measured externally. This process is the origin of the so-called photovoltaic effect. We note that the voltage generated may be detected directly, and that no bias voltage nor ballast resistor is required.

It is also possible to use a *p–n* junction to detect light if one does apply a bias voltage in the reverse direction. By reverse direction, we mean the direction of low current flow, that is, with the positive voltage applied to the *n*-type material. A *p–n* junction operated with bias voltage is termed a photodiode. The characteristics of a photodiode are shown in Figure 5-6. The curve labeled "dark" represents conditions in the absence of light; it displays familiar rectification characteristics. The other

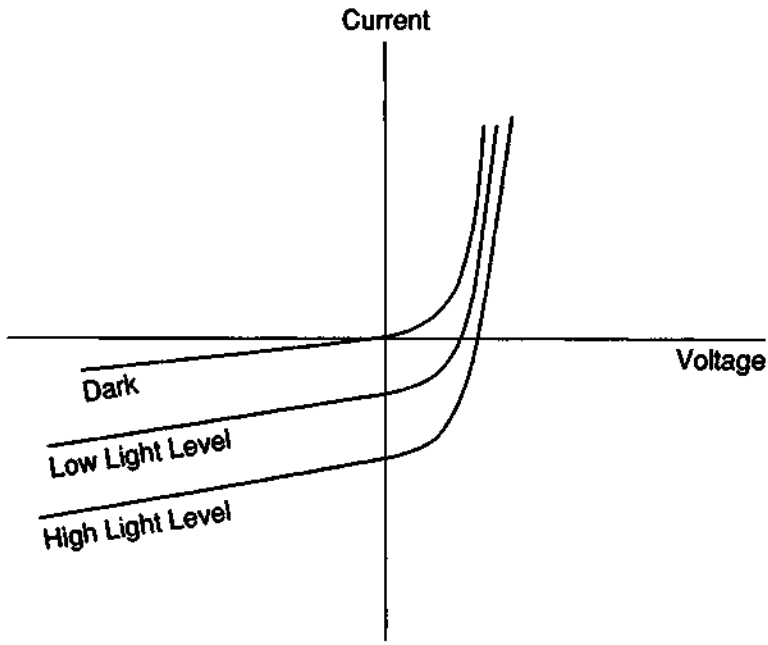


Figure 5-6 Current–voltage characteristics for photovoltaic detectors and photodiodes.

curves show the current–voltage characteristics when the device is illuminated at different light levels. The characteristics of a photovoltaic detector, with zero applied voltage, are represented by the intersections of the different curves with the vertical axis. Even without any applied voltage, an external current is generated by incident light.

The photodiode detector is operated in the lower left quadrant of this figure. The current that may be drawn through an external load resistor increases with increasing light level. In practice, one generally measures the voltage drop appearing across the resistor.

To increase the frequency response of photodiodes, a type called the PIN photodiode has been developed. This device has a layer of nearly intrinsic material bounded on one side by a relatively thin layer of highly doped p-type semiconductor, and on the other side by a relatively thick layer of n-type semiconductor. A sufficiently large reverse bias voltage is applied so that the depletion layer, from which free carriers are swept out, spreads to occupy the entire volume of intrinsic material. This volume then has a high and nearly constant electric field. It is called the depletion region because all mobile charges have been removed. Light absorbed in the intrinsic region produces free electron–hole pairs, provided that the photon energy is high enough. These carriers are swept across the region with high velocity and are collected in the heavily doped regions. The frequency response of such PIN photodiodes can be very high, of the order of 10^{10} Hz. This is higher than the frequency response of p–n junctions without the intrinsic region.

A variety of photodiode structures are available. No single photodiode structure can best meet all system requirements. Therefore, a number of different types have been developed. These include the planar diffused photodiode, shown in Figure 5-7a, and the Schottky photodiode, shown in Figure 5-7b. The planar diffused photodiode is formed by growing a layer of oxide over a slice of high-resistivity silicon, etching a hole in the oxide, and diffusing boron into the silicon through the hole. This structure leads to devices with high breakdown voltage and low leakage current.

The Schottky barrier photodiode uses a junction between a metallic layer and a semiconductor. If the metal and the semiconductor have work functions related in the proper way, this can be a rectifying barrier. The junction is formed by oxidation of the silicon surface, etching of a hole in the oxide, and then evaporation of a thin transparent and conducting gold layer. The insulation guard rings serve to reduce the leakage current through the devices.

Silicon photodiodes respond over the approximate spectral range of 0.4–1.1 μm, covering the visible and near infrared regions. The spectral responsivity of typical commercial silicon photodiodes is shown in Figure 5-8. The responsivity reaches a peak value around 0.7 A/W near 900 nm, decreasing at longer and shorter wavelengths. Optional models provide somewhat extended coverage in the infrared or ultraviolet regions. Silicon photodiodes are useful for detection of many of the most common laser wavelengths, including argon, He–Ne, AlGaAs, and Nd:YAG. As a practical matter, silicon photodiodes have become the detector of choice for many laser applications. They represent well-developed technology and are widely avail-

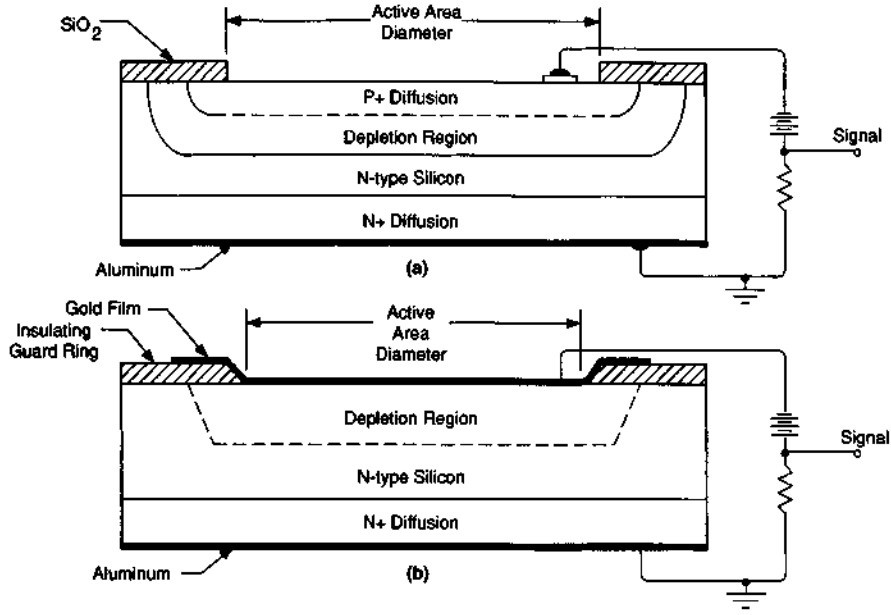


Figure 5-7 Photodiode structures. (a) Planar diffused photodiode. (b) Schottky photodiode.

able. They are the most widely used type of laser detector for lasers operating in the visible and near infrared portions of the spectrum.

Other types of photodiodes are available for lasers operating in other wavelength regions. Photodiodes have been fabricated from many other materials. Different detector materials are useful in different spectral regions. Figure 5-9 shows the spectral D^* (or detectivity) for a number of commercially available detectors, including photovoltaic detectors and photoconductive detectors, which will be discussed later.

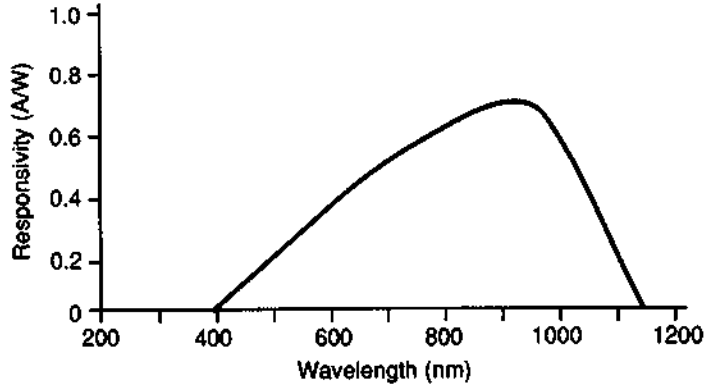


Figure 5-8 Responsivity as a function of wavelength for typical silicon photodiodes.

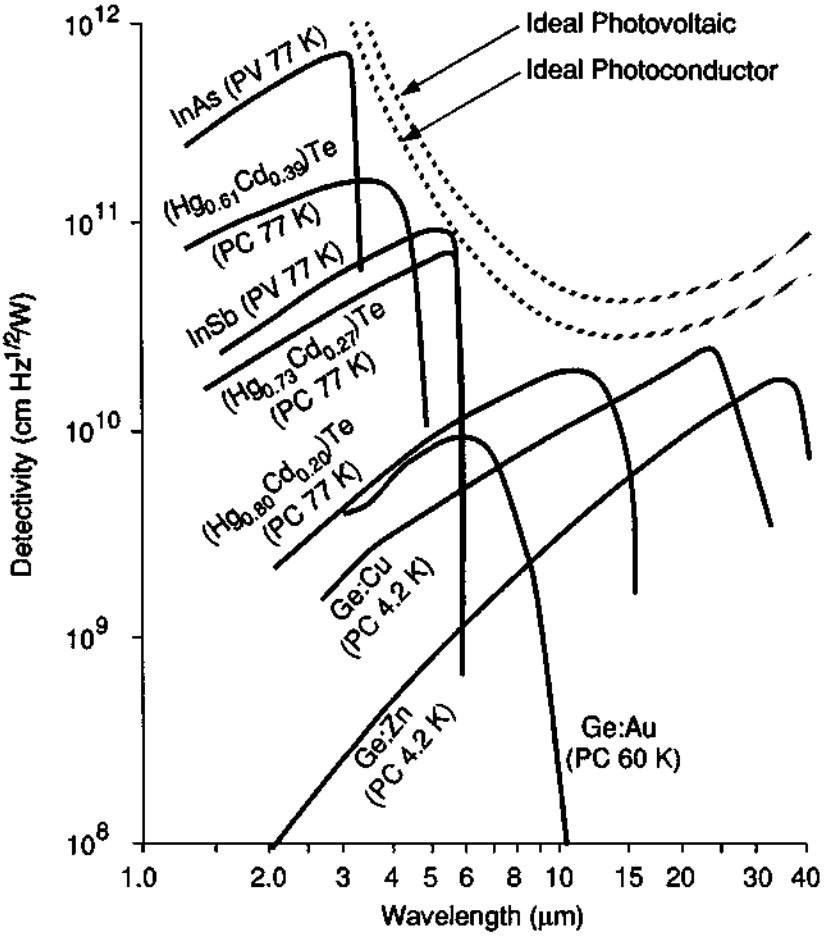


Figure 5-9 Detectivity as a function of wavelength for a number of different types of photodetectors. The temperature of operation is indicated. Photovoltaic detectors are denoted PV; photoconductive detectors are denoted PC. The curves for ideal photodetectors assume a 2π steradian field of view and a 295 K background temperature.

The choice of detector will depend on the wavelength region that is desired. For example, for a laser operating at 5 μm, an indium antimonide photovoltaic detector would be suitable.

The shape of the curves in Figure 5-9 is characteristic of photon detectors. One photon produces one electron–hole pair in the material, so long as the photon energy is high enough. Absorption of one photon then gives a constant response, independent of wavelength (provided that the wavelength lies within the range of spectral sensitivity of the detector). One photon of ultraviolet light and one photon of infrared light each produce the same result, even though they have different energy. For constant photon arrival rate, as wavelength increases, the incident power de-

creases, but the response remains the same. Therefore, the value of D^* increases, reaching a maximum at the cutoff wavelength, which is equal to the Planck's constant times the velocity of light divided by the bandgap of the material. At longer wavelengths, the detectivity decreases rapidly because the photons do not have enough energy to excite electrons into the conduction band.

Figure 5-9 also indicates the detectivity for "ideal" detectors, that is, detects the performance of which is limited only by fluctuations in the background of incident radiation, and that do not contribute noise themselves. Available detectors approach the ideal performance limits fairly closely.

Another variation of the photodiode is the avalanche photodiode. The avalanche photodiode offers the possibility of internal gain; it is sometimes referred to as a "solid state photomultiplier." The most widely used material for avalanche photodiodes is silicon, but these photodiodes have been fabricated from other materials, such as germanium. An avalanche photodiode has a diffused p–n junction, with surface contouring to permit high reverse bias voltage without surface breakdown. A large internal electric field leads to multiplication of the number of charge carriers through ionizing collisions. The signal is thus increased, to a value perhaps 100–200 times as great as that of a nonavalanche device. The detectivity is also increased, provided that the limiting noise is not from background radiation. Avalanche photodiodes cost more than conventional photodiodes, and they require temperature compensation circuits to maintain the optimum bias, but they present an attractive choice when high performance is required.

Silicon and other photodiodes have been configured as power meters, which are calibrated so that the detector output may be presented on a display as the laser power. These devices may be used directly to measure the power of a continuous laser or the average power of a repetitively pulsed laser operating at a reasonably high pulse repetition rate. Some commercial units can measure powers down to the nanowatt regime. Because the photodiode response varies with wavelength, the manufacturer usually supplies a calibration graph to allow conversion to wavelengths other than that for which it was calibrated. Alternatively, some units are supplied with filters to compensate for the wavelength variation of the detector response, so that the response of the entire unit will be wavelength-independent, at least over some interval. Such packaged power meters provide a convenient, useful monitor of laser output power.

2. *PHOTOEMISSIVE DETECTORS*

A photoemissive detector employs a cathode coated with a material that emits electrons when light of wavelength shorter than a certain value falls on the surface. The electrons emitted from the surface may be accelerated by a voltage to an anode, where they give rise to a current in an external circuit. These devices must be operated in vacuum. The frequency response is high, often of the order of 10^8 Hz. These detectors are available commercially from several manufacturers. They represent an important class of detectors for laser applications.

Some standardized spectral response curves for photoemissive cathodes are shown in Figure 5-10. The materials have low work functions; that is, incident light may easily cause the surfaces to emit an electron. The cathodes are often mixtures containing alkali metals, such as sodium and potassium. These devices can be used from the ultraviolet to the near infrared. At wavelengths longer than 1.2 μm, no photoemissive response is available. The short-wavelength end of the response curve is set by the nature of the window material used in the tube containing the detector. The user can choose a device with a cathode having maximum response in a selected wavelength region.

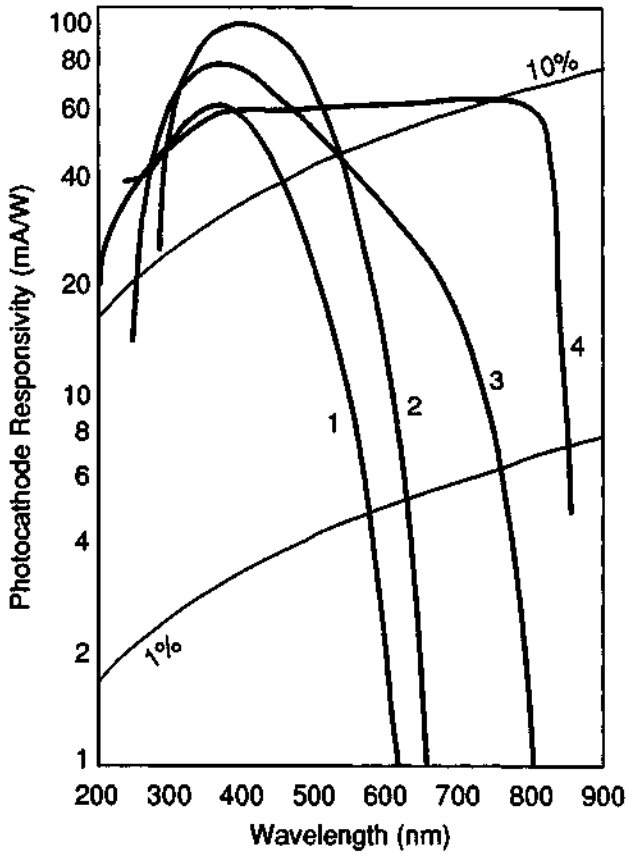


Figure 5-10 Response as a function of wavelength for a number of photoemissive surfaces. Curve 1 is the response of a bialkali type of cathode with a sapphire window; curve 2 is for a different bialkali cathode with a lime glass window; curve 3 is for a multialkali cathode with a lime glass window; and curve 4 is for a GaAs cathode with a 9741 glass window. The curves labeled 1% and 10% denote what the response would be at the indicated value of quantum efficiency.

An important variation of the photoemissive detector is the photomultiplier. This is a device with a photoemissive cathode and a number of secondary emitting stages, called dynodes. The dynodes are arranged so that electrons from each dynode are delivered to the next dynode in the series. Electrons emitted from the cathode are accelerated by an applied voltage to the first dynode, where their impact causes emission of numerous secondary electrons. These electrons are accelerated to the next dynode and generate even more electrons. Finally, electrons from the last dynode are accelerated to the anode and produce a large current pulse in the external circuit. The process has produced a large multiplication factor, so that many (perhaps 10^4 or 10^5) electrons reach the anode for each photon striking the cathode. This high-gain process means that photomultiplier tubes offer the highest available responsivity in the ultraviolet, visible, and near infrared portions of the spectrum. Photomultiplier tubes can in fact detect the arrival of a single photon at the cathode.

3. *PHOTOCONDUCTIVE DETECTORS*

A third class of detectors utilizes the phenomenon of photoconductivity. A semiconductor in thermal equilibrium contains free electrons and holes. The concentration of electrons and holes is changed when light is absorbed by the semiconductor. The light must have photon energy large enough to cause excitation, either by raising electrons across the forbidden bandgap or by activating impurities present within the bandgap. The increased number of charge carriers leads to an increase in the electrical conductivity of the semiconductor. The device is used in a circuit with a bias voltage and a load resistor in series with it. The change in electrical conductivity leads to an increase in the current flowing in the circuit, and hence to a measurable change in the voltage drop across the load resistor.

Photoconductive detectors are most widely used in the infrared spectrum, at wavelengths where photoemissive detectors are not available and the best photodiodes (silicon and germanium) do not operate. There are many different materials used as infrared photoconductive detectors. Typical values of spectral detectivity for some of the most common devices have been presented in Figure 5-9. The exact value of detectivity for a specific photoconductor depends on the operating temperature and on the field of view of the detector. Most infrared photoconductive detectors operate at cryogenic temperatures, which may involve some inconvenience in practical applications.

4. *THERMAL DETECTORS*

Thermal detectors respond to the total energy absorbed, regardless of wavelength. They have no long-wavelength cutoff in their response, as photon detectors do. The value of D^* for a thermal detector is independent of wavelength. Thermal detectors generally do not offer as rapid response as photon detectors, and for laser work they are not often used in the wavelength region in which photon detectors are most effective ($\leq 1.55\ \mu$m). They are often used at longer wavelengths.

Pyroelectric detectors represent one popular form of thermal detector. These detectors respond to the change in electric polarization that occurs in certain classes of crystalline materials as their temperature changes. The change in polarization, called the pyroelectric effect, may be measured as an open circuit voltage or as a short circuit current. The temporal response is fast enough to respond to very short laser pulses. This behavior is in contrast to that of many other thermal detectors, which tend to be slower than photon detectors. Pyroelectric detectors are often used in conjunction with CO_2 lasers.

The calorimeter is another type of common thermal detector. Calorimetric measurements yield a simple determination of the total energy in a laser pulse, but usually they do not respond rapidly enough to follow the pulse shape. Calorimeters designed for laser measurements usually use a blackbody absorber with low thermal mass with temperature measuring devices in contact with the absorber to measure the temperature rise. Knowledge of the thermal mass coupled with measurement of the temperature rise yields the energy in the laser pulse. The temperature measuring devices include thermocouples, bolometers, and thermistors. Bolometers and thermistors respond to the change in electrical resistivity that occurs as temperature rises. Bolometers use metallic elements; thermistors use semiconductor elements.

Many different types of calorimeters have been developed for measuring the total energy in a laser pulse or for integrating the output from a continuous laser. Because the total energy in a laser pulse is often small, the calorimetric techniques are rather delicate. The absorbing medium must be small enough that the absorbed energy may be rapidly distributed throughout the body. It must be thermally isolated from its surroundings so that the energy is not lost.

One form of calorimeter uses a small hollow carbon cone, shaped so that light entering the base of the cone will not be reflected back out of the cone. Such a design acts as a very efficient absorber. Thermistor beads or thermocouples are placed intimately in contact with the cone. The thermistors form one element of a balanced bridge circuit, the output of which is connected to a display or meter. As the cone is heated by a pulse of energy, the resistance of the bridge changes, leading to an imbalance of the bridge and a voltage pulse that activates the display. The pulse decays as the cone cools to the ambient temperature. The magnitude of the voltage pulse gives a measure of the energy in the pulse. In some designs, two identical cones are used to form a conjugate pair in the bridge circuit. This approach allows cancellation of drifts in the ambient temperature.

A calorimeter using a carbon cone or similar design is a simple, useful device for measurement of laser pulse energy. In the range of energy below 1 J or so, an accuracy of a few percent or better should be attainable. The main sources of error in conical calorimeters for pulsed energy measurements are loss of some of the energy by reflection, loss of heat by cooling of the entire system before the heat is distributed uniformly, and imperfect calibration. Calibration using an electrical current pulse applied to the calorimeter element can make the last source of error small. With careful technique, the other sources of error can be held to a few percent.

Calorimeters using absorbing cones or disks with thermocouples to sense the temperature rise have been developed for laser pulses with energy up to hundreds of

joules. When the laser energy becomes high, destructive effects, such as vaporization of the absorbing surface, may limit the usefulness of calorimeters. Because calorimeters require surface absorption of the laser energy, there are limits to the energy that a calorimeter can withstand without damage.

If the response of the calorimeter is fast, it can be used for measurement of power in a continuous laser beam. The temperature of the absorber will reach an equilibrium value dependent on the input power. Such units are available commercially as laser power meters, with different models capable of covering the range from fractions of a milliwatt to 10 kW.

Compared with the power meters based on silicon or other photodiodes, those based on absorbing cones or disks are useful over a wider range of wavelength and do not require use of a compensating factor to adjust for the change in response as the laser wavelength changes. Also, power meters based on these thermal detectors tend to cover a higher range of laser power than do the models based on photodiodes.

Some values of D^* for thermal detectors are shown in Figure 5-11. The values are independent of wavelength. In the visible and near infrared, the values of D^* for thermal detectors tend to be lower than for good photon detectors, but the response does not decrease at long wavelength.

5. *CALIBRATION*

The response of any photodetector in voltage (or current) per unit input of laser power is often taken as the nominal value specified by the manufacturer. For precise work, the detector may have to be calibrated by the user. But accurate absolute measurements of power or energy are difficult. A good calibration requires painstaking work.

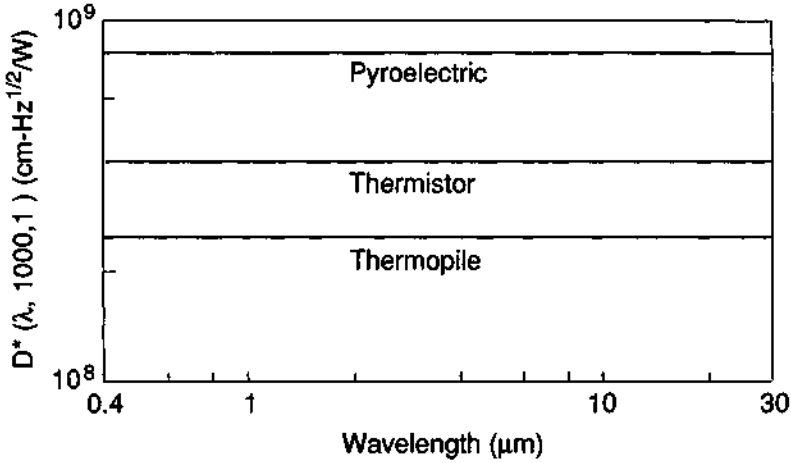


Figure 5-11 Detectivity (D^*) as a function of wavelength for several typical thermal detectors. The temperature of operation is 295 K.

Quantitative measurements of laser output involve several troublesome features. The intense laser output tends to overload and saturate the output of detectors if they are exposed to the full power. Thus, absorbing filters may be used to cut down the input to the detector. A suitable filter avoids saturation of the detector, keeps it in the linear region of its operating characteristics, shields it from unwanted background radiation, and protects it from damage. Many types of attenuating filters have been used, including neutral density filters, semiconductor wafers (like silicon), and liquid filters. Gelatin or glass neutral density filters and semiconductor wafers are subject to damage by high-power laser beams. Liquid filters containing a suitable absorber (for example, an aqueous solution of copper sulfate for ruby lasers) are not very susceptible to permanent damage.

The calibration of filters is a difficult task, because the filters also saturate and become nonlinear when exposed to high irradiance. If a certain attenuation is measured for a filter exposed to low irradiance, the attenuation may be less for a more intense laser beam. Filters may be calibrated by measuring both the incident power and the transmitted power, but the measurement must be done at low enough irradiance so that the filter (and the detector) does not become saturated.

One useful method for attenuating the beam before detection is to allow it to fall normally on a diffusely reflecting massive surface, such as a magnesium oxide block. The goniometric distribution of the reflected light is independent of the azimuthal angle and depends on the angle θ from the normal to the surface in the following simple manner:

$$P_\omega \, d\omega = \mathrm{P}_{\mathrm{tot}} \cos\theta \frac{d\omega}{\pi} \tag{5.4}$$

where P_ω is the power reflected into solid angle $d\omega$ at angle θ from the normal, and P_{tot} is the total power. This relation is called Lambert's cosine law, and a surface that follows this law is called a Lambertian surface. There are many practical surfaces that satisfy this relation approximately. The power that reaches the detector after reflection from such a surface is

$$P_{\mathrm{detector}} = P_{\mathrm{tot}} \cos\theta \frac{A_{\mathrm{d}}}{\pi D^2} \tag{5.5}$$

where A_{d} is the area of the detector (or its projection on a plane perpendicular to the line from the target to the detector) and D is the distance from the reflector to the detector. This approximation is valid when D is much larger than the detector dimensions and the transverse dimension of the laser beam. With a Lambertian reflector, the power incident on the photosurface may be adjusted simply in a known way by changing the distance D. The beam may be spread over a large enough area on the Lambertian surface so that the surface is not damaged. The distance D is made large enough to ensure that the detector is not saturated. The measurement of the power received by the detector, plus some geometric parameters, gives the total beam power.

A thin-membrane beam splitter, called a pellicle, is also used for attenuating laser beams. A pellicle inserted in the laser beam at an angle about 45° to the beam re-

flects about 8 percent and is not easily damaged. If one uses several pellicles in series, with each one reflecting 8 percent of the light from the previous pellicle, one can easily reduce the laser power to a tolerable level.

One widely used calibration method involves measurement of the total energy in the laser beam (with a calorimetric energy meter) at the same time that the detector response is determined. The temporal history of the energy delivery is known from the shape of the detector output. Because the power integrated over time must equal the total energy, the detector calibration is obtained in terms of laser power per unit of detector response.

One may use a calorimeter to calibrate a detector, which is then used to monitor the laser output from one pulse to another. A small fraction of the laser beam is diverted by a beam splitter to the detector, while the remainder of the laser energy is delivered to a calibrated calorimeter. The total energy arriving at the calorimeter is determined. The detector output gives the pulse shape. Then, numerical integration yields the calibration of the response of the detector relative to the calorimeter. Finally, the calorimeter is removed, and the beam is used for the desired application, while the detector acts as a pulse-to-pulse monitor.

Electrical calibration of laser power meters has also become common. The absorbing element is heated by an electrical resistance heater. The electrical power dissipation is determined from electrical measurements. The measured response of the instrument to the known electrical input provides the calibration. It is assumed that the deposition of a given amount of energy in the absorber provides the same response, independent of whether the energy was radiant or electrical.

The difficulty of accurate measurement of radiant power on an absolute basis is well known. Different workers attempting the same measurement often obtain substantially different results. This fact emphasizes the need for care in the calibration of laser detectors.

F. Modulators

We begin the discussion of three types of devices often used with lasers (modulators, deflectors, and Q-switches), each of which can be implemented using either electrooptic or acoustooptic effects. In this section on modulators, we shall describe these phenomena and then show how different devices may be configured with each physical effect.

Amplitude or frequency modulation of laser output is important for many laser applications. Much effort has been devoted to development of methods for high-frequency modulation. The most successful wide-bandwidth systems have employed electrooptic or acoustooptic principles.

1. ELECTROOPTIC MODULATION

The operation of an electrooptic element is shown schematically in Figure 5-12. Polarized light is incident on the modulator. The light may be polarized originally or a

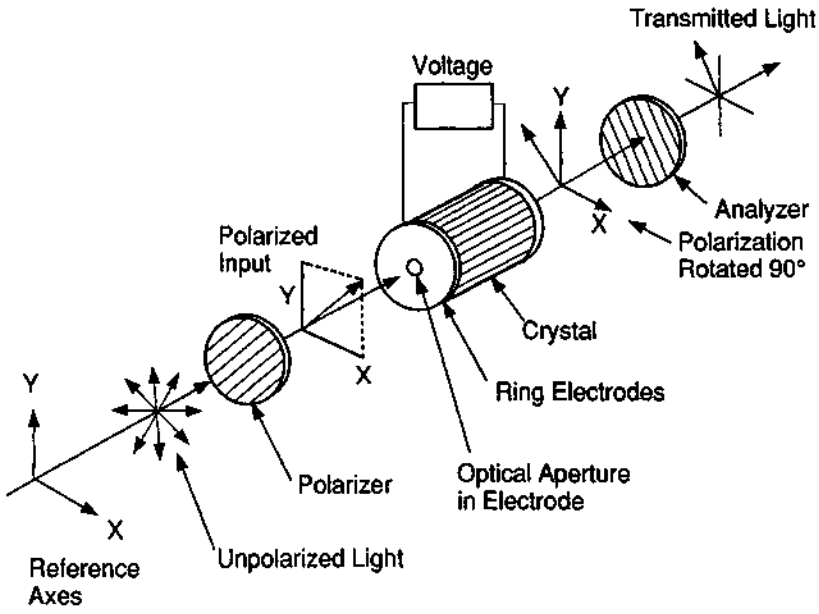


Figure 5-12 Operation of a modulator based on the electrooptic effect. In this configuration, the voltage is applied parallel to the direction of light propagation.

polarizer may be inserted, as shown. The analyzer, oriented at 90° to the polarizer, prevents any light from being transmitted, when no voltage is applied to the electrooptic material. When the correct voltage is applied to the switch, the direction of the polarization is rotated by 90°. Then the light will pass through the analyzer.

Two types of electrooptic effect have been used: the Kerr electrooptic effect, which is shown by liquids such as nitrobenzene, and the Pockel's electrooptic effect, shown by crystalline materials such as ammonium dihydrogen phosphate or lithium niobate. Some early electrooptic devices used nitrobenzene, but the liquid tends to polymerize in the presence of the intense laser light. Modern electrooptic modulators use the Pockel's effect.

In Figure 5-12 the direction of polarization is shown at 45° to the vertical direction. The polarization vector is composed of two perpendicular components of equal intensity, one vertical and one horizontal. The crystalline element is oriented with its axes in a specified orientation (which depends on the crystalline symmetry of the particular material). The applied voltage induces birefringence in the crystal, so that the two components of polarization travel with different velocities inside the crystal. This induced birefringence is the basis of the electrooptic effect.

The two components travel in the same direction through the crystal and do not become physically separated. But the two components, in phase as they enter the crystal, emerge with different phases. As they traverse the crystal, they accumulate a phase difference, which depends on the distance traveled and on the applied voltage.

When the beams emerge from the crystal, the polarization of the combined single beam depends on the accumulated phase difference. If the phase difference is one-half wavelength, the polarization is rotated by 90° from its original direction. This by itself does not change the intensity of the beam. But with the analyzer, the transmission of the entire system varies, according to

$$T = T_0 \sin^2\left(\frac{\pi \Delta n L}{\lambda}\right) \tag{5.6}$$

where T is the transmission, T_0 the intrinsic transmission of the assembly, taking into account all the losses, Δn the birefringence (that is, the difference in refractive index for the two polarizations), L the length of the crystal, and λ the wavelength of the light. The birefringence is an increasing function of the applied voltage, so that the transmission of the device will be an oscillatory function of applied voltage. The maximum transmission occurs when

$$\Delta n = \frac{\lambda}{2L} \tag{5.7}$$

This occurs at a voltage called the half-wave voltage, denoted $V_{1/2}$. The half-wave voltage for a particular material increases with the wavelength. Thus, in the infrared the required voltage is higher than in the visible. This factor can limit the application of electrooptic modulators in the infrared.

One important performance parameter for electrooptic modulators is the extinction ratio, defined as the ratio of the transmission when the device is fully open, to the transmission when the device is fully closed. In practice, there is always some light leakage, so that the minimum transmission never reaches zero. The extinction ratio determines the maximum contrast that may be obtained in a system that uses the modulator. A high extinction ratio thus is desirable. Commercial electrooptic modulators can have extinction ratios in excess of 1000.

Electrooptic modulators may be fabricated with different physical forms. In one form, voltage is applied parallel to the light propagation, as was shown in Figure 5-12. One uses transparent electrodes or electrodes with central apertures. This is called a longitudinal form.

In another form, metal electrodes are on the sides of the crystal (which has a square cross section) and the voltage is perpendicular to the light propagation. This is called a transverse modulator.

Longitudinal and transverse modulators have different applications. Longitudinal modulators may have large apertures, but the half-wave voltage will be high, in the kilovolt range. Longitudinal modulators are often used with conventional light sources. Because it is difficult to vary the high voltages rapidly, the frequency response of longitudinal modulators tends to be low. Applications for which longitudinal modulators are suited include relatively low-frequency modulation of laser beams or use with nonlaser sources.

In transverse modulators, the voltage is applied perpendicular to the light propagation. The phase retardation for a given applied voltage may be increased simply

by making the crystal longer. The half-wave voltage thus may be much lower for transverse modulators. It follows that the frequency response may be higher. Transverse modulators are thus suited for fast, broadband applications. A problem with transverse modulators is the relatively small aperture that they permit, because the crystal may be a long, thin parallelepiped. This physical configuration means that transverse modulators are best suited for use with narrow, well-defined laser beams. In addition, the extinction ratios available with transverse modulators tend to be lower than with longitudinal modulators. Applications of transverse modulators include broadband optical communications, display and printing systems, and fast image and signal recorders.

Table 5-2 lists materials and performance characteristics for some commercially available electrooptic light modulators. Most of the materials are suited for use in the visible and near infrared portions of the spectrum, but cadmium telluride is useful as a modulator for CO_2 lasers.

2. *ACOUSTOOPTIC MODULATORS*

A different approach to modulator technology uses the interaction between light and sound waves to produce changes in optical intensity, phase, frequency, and direction of propagation. Acoustooptic modulators are based on the diffraction of light by a column of sound in a suitable interaction medium. A piezoelectric transducer is attached to the medium and is excited by a radiofrequency driver to launch a traveling acoustic wave. The elastooptic properties of the medium respond to the acoustic wave so as to produce a periodic variation of the index of refraction. A light beam incident on this disturbance is partially deflected in much the same way that light is deflected by a diffraction grating. The operation is shown in Figure 5-13. The alternate compressions and rarefactions associated with the sound wave form a grating, which diffracts the incident light beam. No light is deflected unless the acoustic wave is present.

For a material with a fixed acoustic velocity, the acoustic wavelength or grating spacing is a function of the radiofrequency drive signal; the acoustic wavelength controls the angle of deflection. The amplitude of the disturbance, a function of the

Table 5-2 **Electrooptic Modulator Materials**

Material	Abbreviation	Chemical Formula	Transmission Range (μm)	Bandwidth (MHz)
Ammonium dihydrogen phosphate	ADP	$NH_4H_2PO_4$	0.3–1.2	to 500
Ammonium dideuterium phosphate	AD*P	$NH_4D_2PO_4$	0.3–1.2	to 500
Cadmium telluride	—	CdTe	2–16	to 1000
Lithiun niobate	LN	$LiNbO_3$	0.5–2	to 8000
Lithium tantalate	—	$LiTaO_3$	0.4–1.1	to 1000
Potassium dideuterium phosphate	KD*P	KD_2PO_4	0.3–1.1	to 350

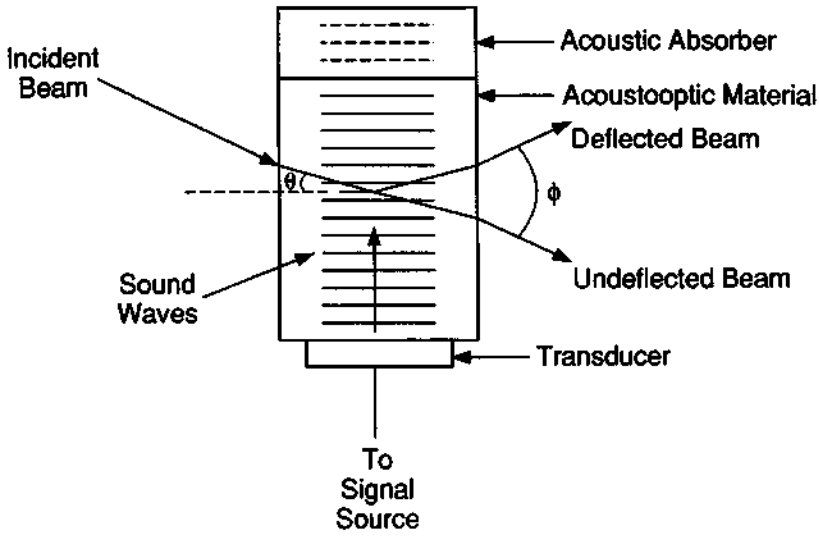


Figure 5-13 Operation of an acoustooptic light beam modulator or deflector, defining the Bragg angle θ and deflection angle ϕ used in the text.

radiofrequency power applied to the transducer, controls the fraction of the light that is deflected. Thus, the power to the transducer controls the intensity of the deflected light. Modulation of the light beam is achieved by maintaining a constant radiofrequency, allowing only the deflected beam to emerge from the modulator, and modulating the power to the transducer. Thus, the modulator will be in its off state when no acoustic power is applied, and it will be switched to its transmissive state by the presence of acoustic power.

The transmission T of an acoustooptic modulator is

$$T = T_0 \sin^2\left[\frac{\pi(M_2 PL/2H)^{0.5}}{\lambda \cos\theta}\right] \tag{5.8}$$

where P is the acoustic power supplied to the medium, L is the length of the medium (length of the region in which the light wave interacts with the acoustic wave), H is the width of the medium (width across which the sound wave travels), and the inherent transmission T_0 is a function of reflective and absorptive losses in the device. The Bragg angle θ is defined by

$$\sin\theta = \lambda/2n\Lambda \tag{5.9}$$

where n is the index of refraction of the material and Λ is the acoustic wavelength. The quantity M_2 is a figure of merit, a material parameter that indicates the suitability of a particular material for this application. It is defined by

$$M_2 = \frac{n^6 p^2}{\rho v^3} \tag{5.10}$$

where p is the material's photoelastic constant, ρ is its density, and v is the velocity of sound in the material. This figure of merit relates the diffraction efficiency to the acoustic power for a given device. It is useful for specifying materials to yield high efficiency, but in practical devices, bandwidth is also important, so that specifying a high value of M_2 is not by itself sufficient to ensure desirable characteristics.

The design and performance of acoustooptic beam modulators have several limitations. The transducer and acoustooptic medium must be carefully designed to provide maximum light intensity in a single diffracted beam, when the modulator is open. The transit time of the acoustic beam across the diameter of the light beam imposes a limitation on the rise time of the switching and, therefore, limits the modulation bandwidth. The acoustic wave travels with a finite velocity, and the light beam cannot be switched fully on or fully off until the acoustic wave has traveled all the way across the light beam. Therefore, to increase bandwidth, one focuses the light beam to a small diameter at the position of the interaction so as to minimize the transit time. Frequently, the diameter to which the beam may be focused is the ultimate limitation for the bandwidth. If the laser beam has high power, it cannot be focused in the acoustooptic medium without damage.

Acoustooptic light beam modulators have a number of important desirable features. The electrical power required to excite the acoustic wave may be small, less than 1 W in some cases. High extinction ratios are obtained easily, because no light emerges in the direction of the diffracted beam when the device is off. A large fraction, up to 90 percent for some commercial models, of the incident light may be diffracted into the transmitted beam. Acoustooptic devices may be compact and may offer an advantage for systems where size and weight are important. Compared with electrooptic modulators, they tend to have lower bandwidth, but they do not require high voltage.

Table 5-3 presents the characteristics of some materials used in commercially available acoustooptic modulators. Several materials are available for use in the visible and near infrared regions, and one material (germanium) is useful in the far infrared. The table also presents some values for the figure of merit M_2, at a wave-

Table 5-3 **Acoustooptic Modulator Materials**

Material	Chemical formula	Spectral range (μm)	Figure of merit, M_2 (10^{-15} m^2/W)	Bandwidth (MHz)	Typical drive power (W)
Fused silica/quartz	SiO_2	0.3–1.5	1.6	to 20	6
Gallium arsenide	GaAs	1.0–1.6	104	to 350	1
Gallium phosphide	GaP	0.59–1.0	45	to 1000	50
Germanium	Ge	2.5–15	840	to 5	50
Lead molybdate	$PbMoO_4$	0.4–1.2	50	to 50	1–2
Tellurium dioxide	TeO_2	0.4–1.5	35	to 300	1–2

length of 633 nm, except for GaAs (1530 nm) and Ge (10.6 μm). The factors that enter into the definition of M_2 vary with crystal orientation; the values in the table are for crystals oriented so as to maximize M_2.

G. Light Beam Deflectors

Light beam deflectors, sometimes called scanners, form an important class of accessory, which is required for applications such as display, optical processing, and optical data storage. Three main methods of light beam deflection are commonly used: mechanical, electrooptic, and acoustooptic. The capabilities of present systems are summarized in Figure 5-14. Light beam deflection is measured in terms of "resolvable spots," rather than in terms of the absolute angle of deflection. One resolvable spot means that the beam is deflected by an angle equal to its own angular spread. The access time is the random access time to acquire one specific spot or resolution element. The figure indicates that mechanical methods give the largest number of resolvable spots, but are slowest, whereas electrooptic devices are fastest, but have the smallest number of resolvable spots.

The figure characterizes the speed in terms of random access time to reach any desired spot. This is a useful parameter for applications like a random access memory. Another possible method of characterizing speed would be the spot scan rate

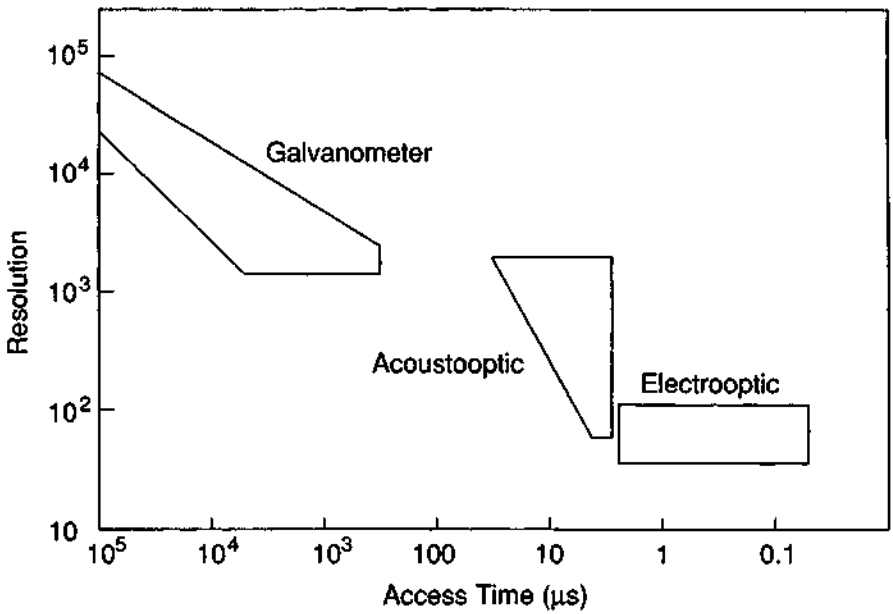


Figure 5-14 Trade-off between resolution (number of resolvable spots) and access time for various types of light beam deflectors.

during a sequential scan. This would be useful for applications like imaging or projection display. Again, the mechanical systems are slowest and the electrooptic fastest, but the difference is less pronounced than for the random access time.

We shall now describe the three methods for light beam deflection in more detail.

1. MECHANICAL DEFLECTORS

Mechanical deflectors consist of mirrors or prisms that are spinning on shafts or rotated by galvanometric or piezoelectric means. Mechanical deflectors offer several advantages, including

- Large deflection angle or large number of resolvable spots
- Low loss of power in deflected beam
- Relatively low drive power
- Wide variety in scanning mode

In general, the deflection accuracy and scanning speed are both low because of the inertia of the moving mass. Figure 5-14 shows that mechanical deflectors produce large deflection angles with relatively low speed, as compared with nonmechanical methods. Thus, mechanical techniques are good in applications where many resolvable spots are required, and electrooptic or acoustooptic techniques are better when fast access time is desired.

Three main types of mechanical deflectors are commonly encountered: rotating mirrors, galvanometer mirrors, and piezoelectric deflectors. Rotating mirror deflectors often use a polygon-shaped mirror mounted on a rapidly rotating shaft. The sides of the polygon are polished to a mirror finish. The laser light strikes a side of the polygon and is scanned in a line as the mirror rotates. When a new side of the polygon rotates into the beam, the line is rescanned, starting at the beginning.

If the polygon has N sides, incident laser light reflected off each side is deflected through an angle of $360°/N$. If the motor speed is w revolutions per second. the scan rate is Nw lines per second. Deflectors of this type have used motors with speeds as high as 2000 revolutions per second. This can lead to scan rates of hundreds of millions of resolvable spots per second along the scan line. But the access time (the time to send the beam to any specified spot) is long, because of the sequential nature of the scan. One must scan through the entire sequence of spots until the desired spot is reached. Motor stability and power consumption at high rotation rates are also issues.

Scanners of this type are used in systems where a raster scan is desired to form an image, as in projection displays. There must be an additional device to offset the scan lines slightly between successive lines in order to form the image.

Galvanometer deflectors use a small mirror mounted on a current-carrying coil between the poles of a permanent magnet. When a pulse of electrical current is applied to the coil, it rotates in the field of the magnet. This is the same principle that galvanometers employ. The mirror may be rotated to any position within its range simply by varying the current.

Such deflectors offer several advantages. Compared with the rotating polygon, they do not require a sequential scan; they can go directly to any desired angular position. Compared with the piezoelectric deflectors, to be discussed, they can achieve larger deflection angles and larger numbers of resolvable spots. Galvanometer deflectors represent a mature, well-engineered technology and offer good performance over a wide range of operating parameters. Two-dimensional scans are easily obtained by using two galvanometers mounted at right angles, as illustrated in Figure 5-15. Such an arrangement allows easy positioning of the laser beam at any point on a surface, for example, the surface of a workpiece in a material processing application. In many applications, it has become preferable to move the laser beam over a surface rather than moving the surface under a fixed beam.

One may choose a galvanometer deflector from many commercial models. The choice involves a trade-off between speed and the number of resolvable spots. To have many resolvable spots, one must make the mirror relatively large, so as to reduce the angular spread of the beam set by diffraction. But a large mirror is heavier and slower to drive than a small mirror. Galvanometer deflectors are available in many models, from large devices that can attain tens of thousands of resolvable spots to small units that can reach a given spot within a few hundred microseconds.

Piezoelectric deflectors use mirrors attached to piezoelectric transducers, which impart motion when an electric field is applied. Piezoelectric materials (such as crystalline quartz) exhibit a change in dimension when voltage is applied to them. When a small mirror is attached to such a material, the change of dimension moves the mirror slightly and deflects a laser beam striking the mirror. The existing materials provide relatively small angular deflections, around 0.001 rad (~0.06°) and small

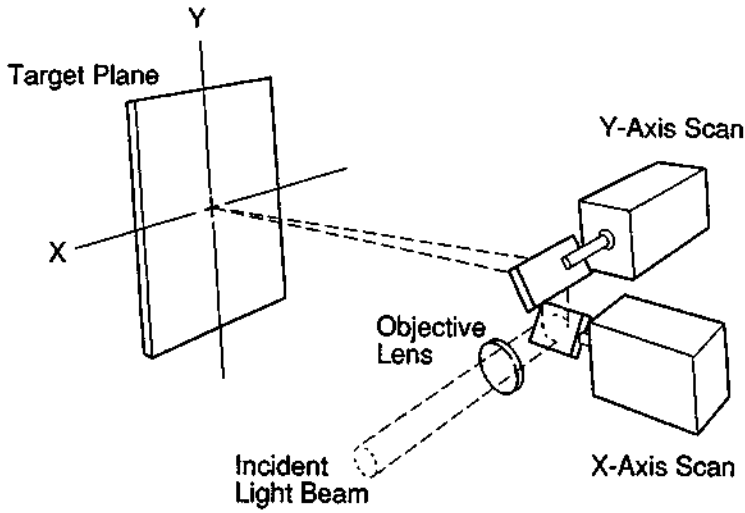


Figure 5-15 Galvanometer arrangement to scan a laser beam over a two-dimensional surface.

numbers of resolvable spots (perhaps a few tens). Because of this limitation, piezoelectric deflectors have had relatively few applications.

To summarize the use of mechanical deflectors, their main advantage involves the absence of volume interaction of light with the material. The deflector has only a single surface, a well-polished mirror. Thus, distortions and light loss are reduced. Also, the cost of mechanical deflectors is generally lower than for the nonmechanical devices.

At the same time, the accuracy of the deflection and the speed are lower than for nonmechanical deflectors. Mechanical deflectors are best suited for applications in which many resolvable spots are desired, and for which the random access to a particular spot need not be fast.

2. *ELECTROOPTIC DEFLECTORS*

Electrooptic deflectors utilize prisms of electrooptic crystals like lithium niobate. A voltage is applied to the crystal; this changes the index of refraction of the material and, hence, the direction of propagation of the beam traveling through the prism. Figure 5-16 illustrates the operation of such a device.

The deflection angle for a plane wave passing through such a prism is obtained by applying Snell's law at the boundaries of the prism. The number N of resolvable spots is

$$N = \Delta n \frac{(L_a - L_b)}{\epsilon \lambda} \tag{5.11}$$

where Δn is the change in the index of refraction, which is a function of the applied voltage, L_a and L_b are the widths defined in Figure 5-16, ϵ is a factor of the order of unity, and λ is the wavelength.

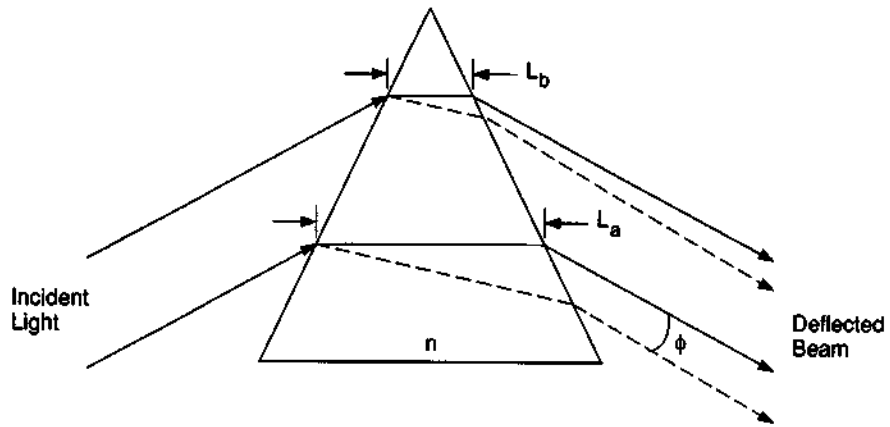


Figure 5-16 Prism setup for an electrooptic light beam deflector element. In the figure, n is the index of refraction of the material and ϕ is the angle of deflection. The figure defines the widths L_a and L_b used in the text.

A method of construction for a useful beam deflector is shown in Figure 5-17. This uses two prisms, one with a positive value of Δn and one with a negative value, when voltage is applied to the electrodes as shown.

Electrooptic deflectors exhibit very fast response and short access time, but they have relatively small numbers of resolvable spots, as Figure 5-14 showed. Other limitations are limited availability of electrooptic crystals with high optical quality throughout a sufficiently large volume and relatively high requirements on operating power. Because of these factors, electrooptic deflectors are used less frequently than galvanometers, rotating mirrors, or acoustooptic devices. Most of the work that has been done on electrooptic deflectors has been experimental; there are relatively few commercial models available.

3. *ACOUSTOOPTIC DEFLECTORS*

Acoustic waves propagating from a flat piezoelectric transducer into a crystal form almost planar wavefronts in the crystal. Light rays passing through the crystal approximately parallel to the acoustic wavefronts are diffracted by the phase grating formed by the acoustic waves. The diffraction satisfies the Bragg condition, familiar in the diffraction of x-rays from planes of atoms in a crystalline lattice. The Bragg angle has been defined in Equation (5.9).

The operation of acoustooptic devices has been illustrated in Figure 5-13 in connection with acoustooptic modulators. When an acoustooptic device is used as a modulator, one observes the diffracted beam at a fixed angle and varies the acoustic

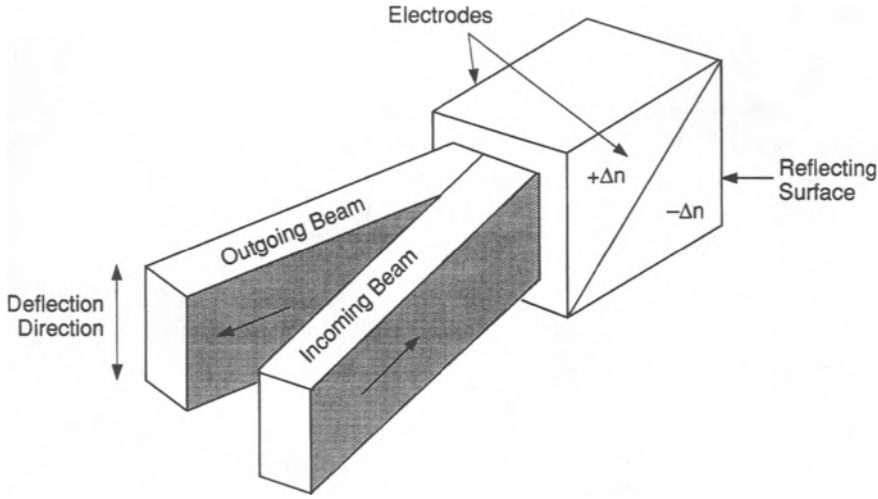


Figure 5-17 A possible construction for an electrooptic light beam deflector, using two prisms, one of which has a positive value of Δn, the induced change of index of refraction, and the other of which has a negative value.

power to change the diffraction efficiency. When one uses an acoustooptic device as a deflector, one keeps the acoustic power constant but varies the acoustic wavelength so as to change the Bragg angle. Then ϕ, the angle formed by the undiffracted and diffracted beams, is

$$\phi \approx \sin \phi = \lambda/\Lambda \tag{5.12}$$

where λ is the optical wavelength and Λ is the acoustic wavelength. Changing Λ thus changes the angle at which the beam emerges from the device. Because the acoustic wavelength is much longer than the optical wavelength, the absolute value of the deflection angle is small, often a few milliradians.

The bandwidth is limited by the time it takes for the acoustic wave to propagate across the diameter of the light beam, the same limitation that applies to the acoustooptic modulator. Acoustooptic deflectors offer several advantages:

- Easy control of deflection modes and position
- Simple structure
- Wide variety of uses from a single device
- Larger number of resolvable spots than electrooptic devices
- Faster access than mechanical devices

The capabilities of acoustooptic deflectors are shown in Figure 5-14. Commercial models offer up to 1000 resolvable spots, with access times typically in the range of 5–50 μsec.

Two-dimensional deflection may be obtained by using two deflectors in series, with their directions of deflection at right angles. Acoustooptic deflectors have been employed in applications such as high-frequency scanning and optical signal processing.

H. *Q*-switches

We have already described the operation of a *Q*-switch in Chapter 2, using a mechanical switch to illustrate the principle of operation. Different types of *Q*-switches have been developed. As we describe them, we shall use a solid state laser as an illustration, but the basic methods are applicable for other types of lasers also.

The *Q*-switching methods all employ a variable attenuator between the laser rod and the output mirror. During the time that the rod is being pumped to a highly excited state, the attenuator is opaque and laser action is inhibited. When the attenuator is suddenly switched to a transparent condition, the laser rod can "see" the output mirror. The energy stored in the rod can be extracted by the stimulated emission process and emitted through the output mirror in a short, high-power pulse.

The various types of attenuator that can be used define three other classes of *Q*-switches. The three types of variable attenuator that have been used are electrooptic elements, acoustooptic elements, and bleachable dyes.

The use of electrooptic and acoustooptic elements as light modulators has been described earlier. It is apparent that a *Q*-switch is a specialized type of modulator. The operation of an electrooptic *Q*-switch is shown schematically in Figure 5-18,

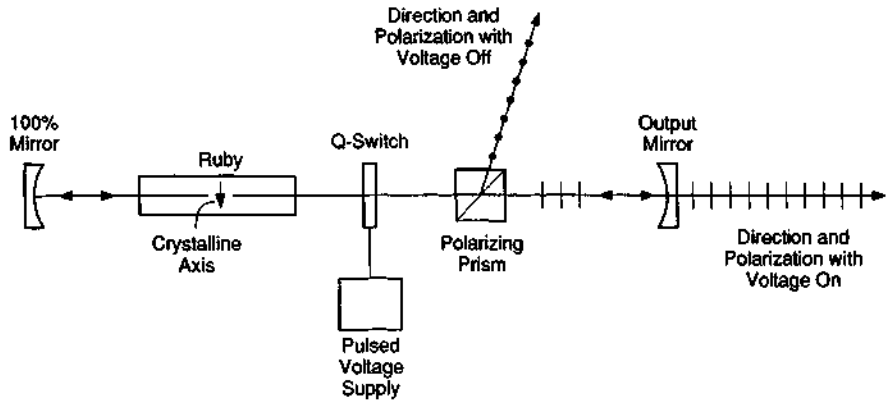


Figure 5-18 Application of an electrooptic Q-switch in a ruby laser.

using a ruby laser as an illustration. The light emitted from the rod in the orientation shown in the figure is polarized primarily in a horizontal direction. The analyzer will prevent the light from reaching the output mirror. When voltage is applied to the electrooptic switch, the plane of polarization of the light is rotated 90°. Then light from the rod passes through the analyzer and reaches the mirror. Electrooptic Q-switches use the Pockel's effect in crystalline materials like potassium dideuterium phosphate. They are extremely fast, with opening times around 1 nsec.

Acoustooptic Q-switches rely on the diffraction grating set up by sound waves in certain materials, as described earlier. The light beam is deflected from its path by diffraction when the sound wave is present. The operation of an acoustooptic Q-switch is shown in Figure 5-19. When the acoustic wave is present, some of the light is diffracted and misses the mirror. The deflection angle, a small fraction of one degree, is exaggerated in the figure. The loss due to diffraction is enough to keep the laser below threshold. When the pumping of the laser is complete, the acoustic wave is switched off. Then all the light reaches the mirror, and a laser pulse builds up rapidly. The most commonly encountered acoustic Q-switch uses fused silica. The opening time is typically around 100 nsec, longer than for the electrooptic device.

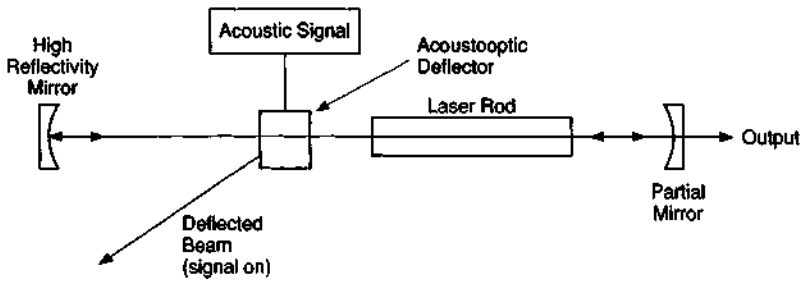


Figure 5-19 Application of an acoustooptic Q-switch in a solid state laser.

Acoustooptic devices have proved to be the most useful type for Nd:YAG lasers. The voltage required for operation of an electrooptic device increases with wavelength, a disadvantage for infrared applications. Bleachable dyes for Q-switching in the infrared are less developed than those for use in the visible. Thus, acoustooptic devices are commonly used for Nd:YAG Q-switching.

Bleachable dye Q-switches are materials that initially absorb light at the laser wavelength, but then bleach and become transparent as the light intensity increases. Figure 5-20 shows the transmission at 694.3 nm of a particular bleachable dye, vanadium phthalocyanine in nitrobenzene solution, as a function of incident power. The switching behavior is apparent. At low power, the dye has about 1 percent transmission and inhibits laser operation. At high power, the dye becomes almost 100 percent transmissive and laser operation can proceed. The physical basis for this behavior arises from saturation of all the absorbing molecules, which are raised from their absorbing ground state to a nonabsorbing excited state. If the light is intense enough, all the molecules are in the excited state, and the material is transparent. Between pulses, the dye molecules relax back to the ground state and become absorbing again.

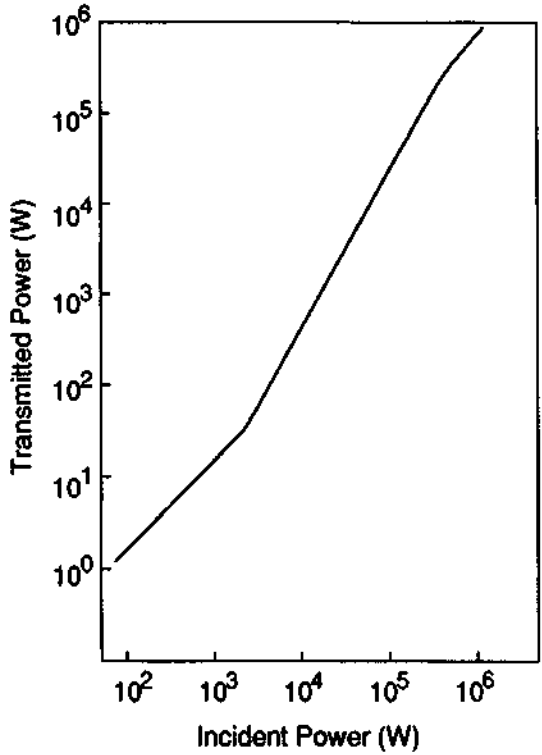


Figure 5-20 Transmission at 694.3 nm of the dye vanadium phthalocyanine in nitrobenzene solution as a function of incident power. The figure illustrates the capability of this dye for bleaching at high light intensity. (From J. A. Armstrong, *J. Appl. Phys.* **36**, 471 (1965).)

Bleachable dyes have been used most often with ruby lasers. A number of dye materials and solvents have been used. The most common is the dye cryptocyanine in methyl alcohol solution. For Nd:glass lasers, the so-called 9740 and 9860 dye solutions have been used.

One disadvantage of the dye Q-switch is that it tends to "open" unevenly. Thus, it produces beams with poor spatial quality.

Table 5-4 compares properties of the various types of Q-switches. The table contains relative comparisons of the characteristics of the Q-switches among themselves, rather than absolute statements. The dynamic loss is the loss introduced by the Q-switch in its off state; high values are desirable. In early lasers, all four types of Q-switches were widely used. Rotating mirror systems have fallen out of use now, primarily because of the slow switching speed. Bleachable dyes offer the simplest and lowest-cost device, but the properties of the laser pulse, particularly the poor spatial profile, are less desirable than for acoustooptic and electrooptic devices. They are used less frequently in modern lasers. In recent times, most Q-switched lasers have employed electrooptic or acoustooptic Q-switches.

I. Nonlinear Optical Elements

The electromagnetic field associated with the passage of a light wave through a material produces a polarization in the material. For simplicity, we consider only the electric polarization vector **P**. In the classical discussion of wave propagation, the polarization is treated as being proportional to the electric field vector **E**. The polarization of a dielectric medium by an applied electric field results from orientation of the molecules under the action of the field. If the electric field varies sinusoidally, the polarization will vary sinusoidally also. When the intensity of the light is low, the polarization is proportional to the electric field strength. But at high intensity, this linear proportionality may no longer be valid. In particular, the intensity available from a laser beam will often exceed the limits of the linear approximation. One then expresses the component P_i of the polarization in the i direction as a series

$$P_i = \epsilon_0 \chi_{ij} E_j + 2d_{ijk} E_j E_k + 4\chi_{ijkl} E_j E_k E_l + \ldots \tag{5.13}$$

where the equation is in tensor notation such that whenever a subscript is repeated in a term, it implies a summation over the subscript, and where E_i is the component of the electric field in the i direction, ϵ_0 is the permittivity of vacuum (a constant),

Table 5-4 **Comparison of Q-Switching Devices**

Type	Cost	Dynamic loss	Pulse-to-pulse stability	Switching speed	Beam profile
Rotating mirror	Low	Very high	Good	Slow	Good
Electrooptic	High	High	Good	Very fast	Good
Acoustooptic	Moderate	Moderate	Good	Medium	Good
Bleachable dye	Very low	Moderate	Fair	Fast	Poor

χ_{ij} is the linear susceptibility of the material, and d_{ijk} and χ_{ijkl} are the second- and third-order nonlinear susceptibilities, respectively. These last two quantities are much smaller than the linear susceptibility. If the electric field is suitably small, the variation of P_i with an applied electric field of the form $E_0 \cos \omega t$ will be of the form $P_0 \cos \omega t$, the classical linear case. But when E_i becomes large enough, the term second term in Equation (5.13) becomes comparable with the first. The second term in the equation will contribute a polarization

$$P_i = 2d_{ijk}E_jE_k \cos^2 \omega t = d_{ijk}E_jE_k(1 + \cos 2\omega t) \tag{5.14}$$

Consider the form $(1 + \cos 2\omega t)$. The first term represents a constant component (or so-called dc component). The second term, which represents an oscillation of electric charge at frequency 2ω, results in generation of a second harmonic. In other words, it leads to radiation of light with frequency twice that of the original incident light. Higher harmonics at higher frequencies can likewise arise from the higher-order terms in the series. They are generally weaker, because the coefficients decrease rapidly in the series in Equation (5.13).

In many materials, the coefficient d_{ijk} will be identically zero. For d_{ijk} to be other than zero, the material must be crystalline, with a crystalline cell that does not have a center of symmetry. In such materials, the passage of a laser wave at frequency ω and with high intensity will cause polarization at frequency 2ω, the second harmonic of the incident light. This polarization then radiates second harmonic radiation, so that one obtains radiation at a frequency twice that of the incident frequency. Before lasers were available, this phenomenon was not observed. All optics was "linear"; that is, only the first term in Equation (5.13) could be observed, because the electric field of the light wave was too small to produce a measurable contribution from the other terms. When lasers became available, the electric field in the light waves became much larger, and the contribution of the higher-order terms could be observed. Thus, "nonlinear" optics has been made possible by the invention of lasers.

Some commonly used nonlinear optical materials are listed in Table 5-5, along with their properties. The nonlinear coefficient given in the table is the largest element in the tensor d_{ijk}. The smallness of its magnitude suggests why nonlinear optical effects had not been observed with nonlaser sources.

In the early history of lasers, materials like potassium dihydrogen phosphate were the only ones available, and the original work in nonlinear optics was done with them. These materials were not durable and were difficult to work with. More recently, better materials have become available, particularly beta-barium borate (BBO) and potassium titanyl phosphate (KTP). Most modern frequency-doubled lasers use one of these materials. KTP is widely used for doubling Nd:YAG lasers to produce green light, and BBO is frequently used to double visible light to reach the ultraviolet.

The intensity $I_{2\omega}$ of the frequency-doubled radiation after the light has traversed a distance L in the material is

$$I_{2\omega} = \frac{\sin^2[(2\pi/\lambda_1)(n_1 - n_2)L]}{(n_1 - n_2)^2} \tag{5.15}$$

TABLE 5-5 **Nonlinear Optical Materials**

Material	Abbreviation	Chemical formula	Transmission range (μm)	Nonlinear coefficient at 1.06 μm (10^{-12} m/V)
Ammonium dihydrogen phosphate	ADP	$NH_4H_2PO_4$	0.3–1.2	0.74
Ammonium dideuterium phosphate	AD*P	$NH_4D_2PO_4$	0.3–1.2	0.8
Barium sodium niobate	BSN or BANANA	$Ba_2NaNb_5O_{15}$	0.45–4.5	14.6
Beta-barium borate	BBO	β-BaB_2O_4	0.19–2.5	1.6
Lithiun niobate	LN	$LiNbO_3$	0.5–2	5.82
Lithium tantalate	—	$LiTaO_3$	0.4–1.1	1.28
Lithium triborate	LBO	LiB_3O_5	0.16–2.6	1.16
Potassium dihydrogen phosphate	KDP	KH_2PO_4	0.3–1.1	0.61
Potassium dideuterium phosphate	KD*P	KD_2PO_4	0.3–1.1	0.63
Potassium titanyl phosphate	KTP	$KTiOPO_4$	0.19–2.5	6.5

where λ_1 is the incident wavelength and n_1 and n_2 are the indices of refraction at the original and frequency-doubled wavelengths, respectively. If $n_1 = n_2$, the intensity will increase as the square of the distance, at least until the power in the incident beam becomes depleted. But the index of refraction is a function of wavelength, and generally $n_1 \neq n_2$. Then, the original and frequency-doubled waves will get out of phase, and the intensity of the doubled wave will oscillate as the distance increases. It will reach a maximum value when the crystal thickness is

$$L_c = \frac{\left(\dfrac{\lambda_1}{|n_1 - n_2|}\right)}{4} \tag{5.16}$$

This thickness, called the coherence length, will be around 10 μm for typical nonlinear materials. This would severely limit the power that could be converted into second harmonic radiation. This limitation can be avoided by the technique of phase matching. In some birefringent crystals, one can choose combinations of the direction of propagation and directions of polarization of the two beams so that $n_1 = n_2$. Then, the intensity $I_{2\omega}$ can build up without the two waves getting out of phase. Exact matching is accomplished by varying the angular orientation of the crystalline axes with respect to the beams, by varying the temperature, or by both. With phase matching, conversion of the incident light into second harmonic radiation is possible with high conversion efficiency. Commercial frequency-doubled Q-switched Nd:YAG lasers can provide output at 532 nm with one-third to one-half of the energy that would have been available at the fundamental wavelength of 1064 nm.

Second harmonic generation works best with pulsed lasers with high peak power, because the polarization that produces it is proportional to the square of the electric field. Thus, the most common frequency-doubled lasers are Q-switched devices. Many commercial frequency-doubled lasers have become available. The most popular of these is the frequency-doubled Nd:YAG laser, with its output in the green at wavelength 532 nm. Frequency-doubled tunable dye lasers with their original output in the visible can provide tunable output in the ultraviolet.

Frequency doubling is the most common manifestation of nonlinear optics. This brief discussion of nonlinear optics, which greatly abbreviates this extensive subject, has emphasized second harmonic generation. Many other phenomena have been investigated. The third term in Equation (5-13) corresponds to frequency tripling, which leads to an output at 355 nm for a Nd:YAG laser. One can mix two beams of different frequencies ω_1 and ω_2 and obtain outputs at the sum frequency $\omega_1 + \omega_2$ or at the difference frequency $\omega_1 - \omega_2$. In this way, many different wavelengths covering a broad spectral range may be obtained, starting with a few basic wavelengths.

J. Optical Isolators

An optical isolator is a device that allows light to propagate through it in one direction, but not in the opposite direction. Isolators are useful as valves that allow propagation in only one direction. They are used in high-power applications, for which one desires one-way transmission of light. In many cases, backward propagation of light, reflected from a workpiece or other element, might send some light back into the laser and could be damaging. In laser amplifiers especially, it is desired to have one-way transmission. If a small amount of light is reflected back into an amplifier, it will be amplified as it passes through stages with progressively smaller cross sections, so that optical damage would become probable. The use of isolators can prevent this. Isolators are also widely used in optical communications applications to prevent unwanted back reflections.

The most common type of isolator uses a rotation of the polarization of the light by means of the Faraday magnetooptic effect, which occurs in certain materials. A diagram of an isolator based on the Faraday magnetooptic effect is shown in Figure 5-21. Light is incident from the left and has its plane of polarization defined by the first polarizer, shown as vertical in this figure. A magnetic field is applied to the material by means of a coil of wire. If the conditions are correct, the direction of polarization will be rotated by 45° as it emerges from the magnetooptic material, and the light will pass through the second polarizer.

Light incident from the right and passing through the second polarizer will have its direction of polarization at 45° from the vertical. As it passes from right to left through the magnetooptic material, the direction of polarization will be rotated to 90° from the vertical, and the light will not pass through the left polarizer. Thus, the device acts as a one-way valve.

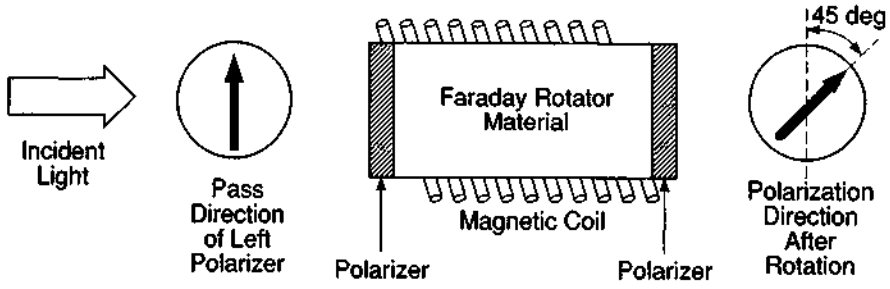


Figure 5-21 An optical isolator using the Faraday megnetooptic effect.

The amount of rotation θ of the direction of polarization is chosen in the positive screw sense along the direction of the applied magnetic field, and it is given by

$$\theta = VBL \tag{5.17}$$

where V is a constant of the material, called the Verdet constant, B is the magnetic field, and L is the length of the magnetooptic material. The length of the material sample and the magnetic field are selected so that the angle of rotation is 45° for the chosen material. Near the red end of the visible spectrum, the magnetooptic materials are glasses with Verdet constants around 0.1–0.2 min/cm-Gauss. The materials are bulk samples, with dimensions of centimeters required to give sufficient rotation for reasonable values of the magnetic field. At the wavelengths of 1.3 and 1.55 μm, materials include yttrium iron garnet and bismuth-substituted rare earth iron garnet, with Verdet constants as high as 800 min/cm-Gauss. Samples of these materials can have submillimeter length and give the desired rotation.

Commercial optical isolators are available with extinction ratio (ratio of the intensity of the light transmitted in the two opposite directions) greater than 40 dB.

K. Raman Shifters

Raman shifting is sometimes employed to obtain a broader range of wavelengths than would otherwise be available from lasers. Raman scattering involves interaction between a photon and a molecule that can be vibrationally excited. The molecule absorbs the photon, with energy hc/λ_0, where λ_0 is the laser wavelength, h is Planck's constant, and c the velocity of light. The molecule then reemits a photon with energy hc/λ_s, where λ_s is a longer wavelength. The difference in energy, $hc(1/\lambda_0 - 1/\lambda_s)$, is used to excite the vibrational energy of the molecule. Because the energy levels of the molecule are quantized, the amount of energy loss is also quantized. This means that the wavelength λ_s has a specific discrete value, which is characteristic of the molecule that did the scattering.

When the molecule absorbs energy from the photon, the process is called Stokes scattering. The scattered light can interact further, undergoing second, third, and so

forth, scattering processes. This can produce a series of Stokes lines at progressively longer wavelengths.

It is also possible for a photon to absorb energy from the molecule if the molecule is in an excited state. This type of scattering is called anti-Stokes scattering and leads to Raman-shifted lines at a shorter wavelength. Usually anti-Stokes lines are less intense than Stokes lines.

The result of a Raman-shifting process is shown in Figure 5-22, which presents the redistribution of energy from a 249-nm KrF laser after Raman scattering in H_2 gas. The figure shows that the original laser energy can be Raman shifted to other wavelengths with reasonable efficiency.

Some commercial models of Raman shifters are available, usually in the form of pressurized gas cells. Different gases lead to wavelength shifts by different amounts. By using Raman shifting, one may obtain a wide variety of wavelengths, in addition to the wavelengths of available lasers. Because Raman shifting occurs with higher efficiency at short wavelengths, it is most often used with ultraviolet lasers, especially excimer lasers, but there are many examples of its use at longer wavelengths, including use in the infrared. When Raman shifting is employed with a tunable laser, it extends the range over which the laser may be tuned.

L. Injection Seeders

Injection seeding involves directing the output of a small, stable, single-frequency laser, called the seed laser, into the laser cavity of a larger laser in order to stabilize

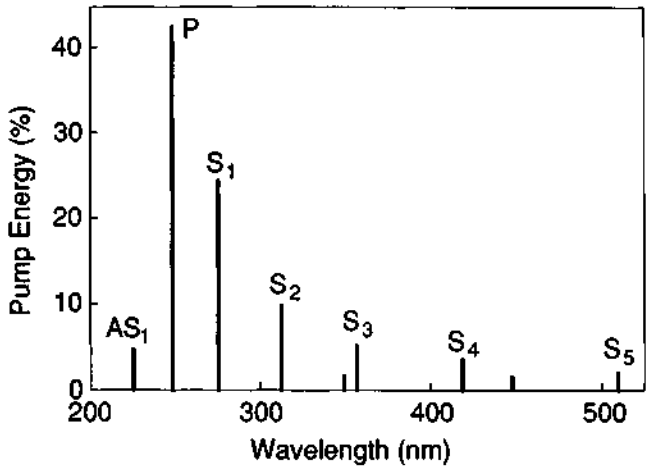


Figure 5-22 Distribution of energy from a KrF laser operating at 249 nm into a number of Raman-shifted lines by means of Raman scattering in hydrogen at a pressure of 20 atm. The original pump laser wavelength is denoted P, and the Stokes and anti-Stokes lines are denoted S and AS, respectively. (From T. R. Loree, R. C. Sze, and D. L. Barker, *Appl. Phys. Lett.* **31**, 37 (1977).)

the output of the larger laser. It is most often used with pulsed lasers, for example, Q-switched lasers. The seed laser is usually a continuous laser. The frequency of the seed laser must lie within the linewidth of the material for the larger laser. Then, when the pulse of the larger laser develops, it occurs at the longitudinal mode that is closest to the frequency of the seed laser. Thus, the output of the pulsed laser will be at a single frequency, rather than in a number of longitudinal modes at different frequencies.

In addition to stabilizing the frequency of the pulsed laser, injection seeding can reduce variations in the output energy from pulse to pulse, improve the beam spatial profile, reduce jitter in the timing of the pulse, and smooth out temporal variations in the pulse. Figure 5-23 shows how the output pulse shape of a Q-switched Nd:YAG laser changes when it is injection seeded by a low-power continuous Nd:YAG laser beam.

Commercial models of injection seeding lasers are available, including models based on semiconductor laser diodes as the seeding source.

M. Beam Profilers

For many applications, it is important to know the shape of the laser beam and the distribution of power within it. Beam profiling instrumentation has developed to provide easy, rapid characterization of laser beams.

There are two main approaches to determining the profile of a beam. The first is to direct the beam into a large-area detector and pass an opaque mask in front of the detector. The mask may take the form of a slit, a knife edge, or a pinhole. The decrease in signal as the mask is scanned in front of the detector is determined, and the intensity profile is derived. The second method is to use an array of detectors, which essentially forms a camera. A camera consisting of an array of semiconductor charge-coupled devices (CCDs) is commonly used.

In either case, the resulting signals are processed, and the beam profile and power distribution are displayed. Commercial beam profile equipment has many software options and display formats. The profiles can be displayed as two-dimensional plots, isometric three-dimensional plots, or as cross-sectional profiles.

The mechanical devices are well suited for use with continuous lasers or lasers with high pulse repetition rate. They can have resolution in the submicrometer regime and thus can be used to determine profiles near a focal point of the beam. The camera devices are suited for single-shot profiling of pulsed beams with low repetition rate. They are also useful for beams with large cross sectional area. The resolution of the camera profilers is defined by the number of detectors in the array. Two-dimensional arrays with hundreds of detectors in each dimension are available.

The spectral response of a beam profiler depends on the detector that is used. Available models can cover the range from the ultraviolet to the far infrared. Beam profiling equipment can be valuable to a laser user for defining beam parameters and for correcting degradation of the beam quality.

Related types of instrumentation are available for determining beam quality in terms of the parameter M^2 defined in Chapter 2. Commercial instrumentation has

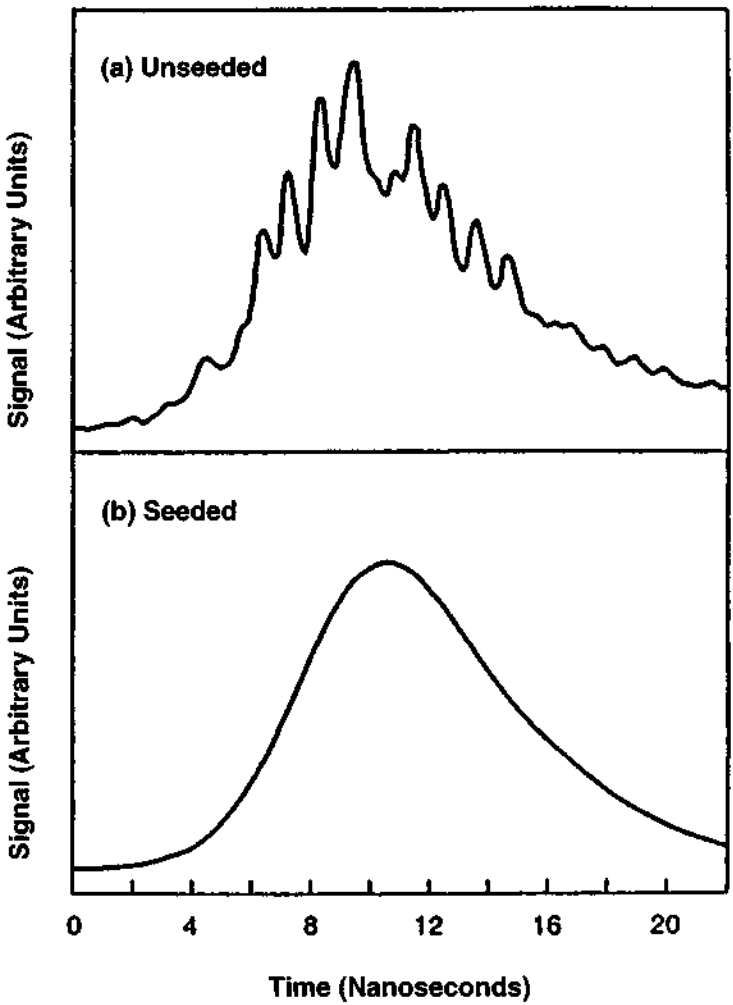


Figure 5-23 Pulse shape of a Q-switched Nd:YAG laser, with (a) and without (b) injection seeding. (From LIGHTWAVE Electronics Corp., *Introduction to Diode-Pumped Solid-State Lasers*, Mountain View, CA, (1993).)

been developed for automatic measurement of this quantity. In one approach, after the measurement head is positioned in the beam, the system uses a servo driven lens to focus the beam on the rear of a rotating drum inside the head. The rotating drum has two knife edges, oriented perpendicular to each other, which block the beam as the drum rotates. This allows determination of the beam diameter, both at the front of the head and at the rear, The second determination is made when the drum has rotated 180° from the position of the first determination. A microprocessor then uses the results of the measurement to calculate the value of M^2. One also can calculate

other relevant parameters, such as the beam divergence, astigmatism, asymmetry, and beam diameter at any position along the beam axis.

N. Optical Tables

Many laser applications require an extremely stable platform. A holographic recording, for example, can be degraded by relative motion between the components of less than the wavelength of light. Tables to provide a stable, vibration-free platform have become a standard accompaniment of many laser applications.

The first stable platforms for lasers were large granite slabs, mounted on some vibration isolating structure, sometimes as simple as partially inflated rubber inner tubes. Now, steel honeycomb tables with sophisticated vibration isolation are more in favor.

The stable platform should have a relatively large mass to reduce the motion caused by low-frequency vibrations. It should also have high rigidity, to reduce relative motion between components on the surface due to high-frequency vibrations. Finally, it should have damping to reduce the amount of vibration that reaches the platform.

Granite slabs offer large mass and high rigidity and can be polished to a very flat surface. They might be chosen for applications in which surface flatness or large mass are important. However, large slabs can become excessively heavy. It is difficult to mount components on granite slabs. Thus, granite slabs are used less often today than steel honeycomb tables.

Steel tables are available from a number of manufacturers, with a variety of constructions. They have two surface plates sandwiched around a lightweight damped honeycomb core. They offer a satisfactory combination of desirable characteristics, including high mass, excellent rigidity, and a high ratio of rigidity to mass. Frequently, the top surface has periodically spaced tapped holes for easy mounting of components. If the steel is ferromagnetic, it is also possible to use magnetic mounts for positioning.

Aluminum has also sometimes been used instead of steel for the top of the table. It is lighter in weight, but it is more prone to deformation than steel.

For low-performance applications, one might also use a single steel plate. However, these plates have reduced rigidity and are subject to bending in response to vibration.

The performance of the structure is frequently expressed in terms of its compliance. Compliance is a quantitative expression of the susceptibility of a structure to motion as a result of an external force. It is equal to the displacement caused by an applied force divided by the magnitude of the force. It has the dimension of length over force, and it is presented as a function of frequency. Figure 5-24 shows an example of the compliance of a steel honeycomb table as a function of frequency. The straight line is an inverse square dependence, which would be the response of a rigid body. At low frequencies, the compliance decreases as the inverse square of the

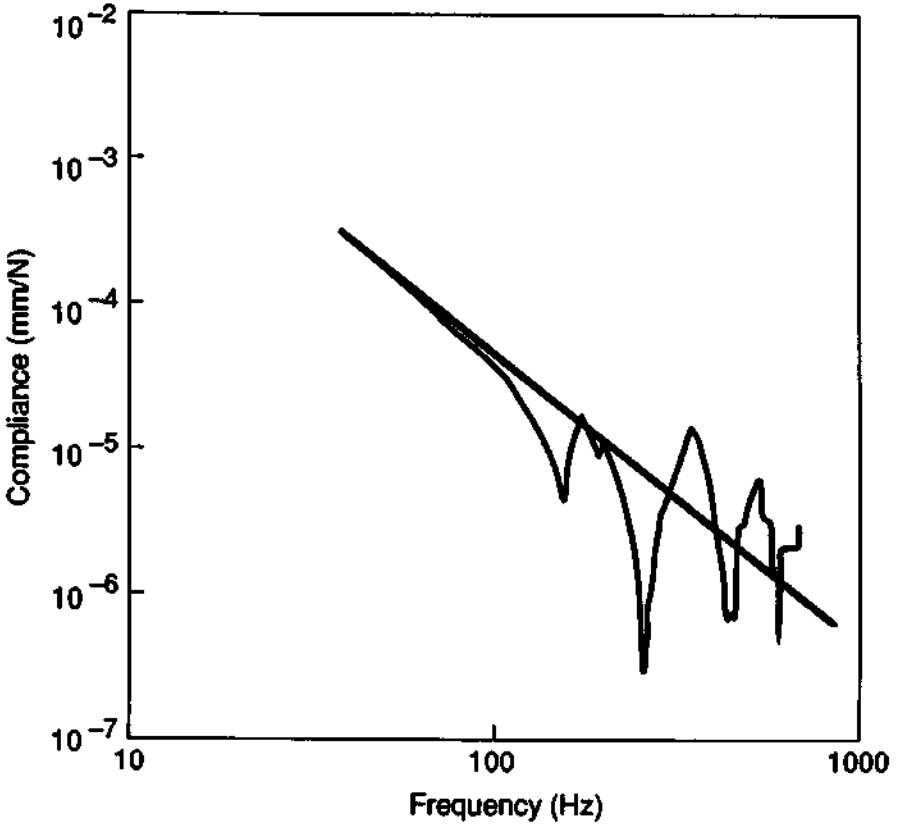


Figure 5-24 Compliance curve for a steel honeycomb table top. (From Newport Corp., *The 1994 Newport Catalog*, Irvine, CA (1994).)

frequency. In other words, it behaves like a rigid body. At higher frequencies, structural vibrations are excited and the platform begins to deform. The peaks in the curve represent resonant frequencies of the structure. The maximum structural bending will be produced by vibrations that have frequencies near the highest peaks in the compliance curve.

The platform is mounted on a damping structure, such as pneumatic isolators, to reduce the amount of vibration reaching the tabletop. The performance of the isolator is characterized by its transmissibility, which is the fraction of the vibration that is transmitted from the environment to the tabletop. The transmissibility is a function of frequency. For many tables, the transmissibility is near unity, or perhaps even greater than unity (meaning that the vibration is amplified) at frequencies around 1 or 2 Hz. The transmissibility decreases rapidly with increasing frequency, and it may be around 0.01 at 10 Hz.

In summary, there are available a variety of different types of stable optical platforms for laser work.

O. Spatial Light Modulators

Spatial light modulators (SLM) are devices that modulate the intensity of light at different positions in a controlled pattern. They can impose a desired spatial pattern of light intensity on a uniform optical beam. They are useful for applications like display, holographic data storage, and optical computing.

In perhaps the simplest form, an SLM could consist of an array of individually addressable shutters, through which a light beam is transmitted. The shutters are opened or closed to impose the desired spatial pattern on the beam. Alternatively, the SLM may operate in reflection, with a spatially varying pattern of reflection controlling the light intensity.

A wide variety of different technologies have been used to fabricate SLMs. These include the following:

Liquid crystal light valves. These are familiar from use in devices like digital watches and laptop computer displays.

Magnetooptic light valves. In these, a magnetic field produces changes in the polarization of light using materials like iron garnet films. The change in polarization is transformed into a change in intensity when the beam is passed through a polarizer.

Acoustooptic devices. These have been described earlier. They may be fabricated in arrays, with each channel controlled by a separate piezoelectric transducer and an oscillator.

Deformable mirrors. In one example, a metallized membrane deposited over an etched cantilever structure and an array of transistors could be deformed by electrostatic charge attracting the cantilevers.

There have been many different types of SLM developed. The few examples just presented will give an idea of some of the common approaches.

We note that an SLM may be controlled by external electric or magnetic fields, as is the case with the examples we have presented. It is also possible to control the operation of an SLM optically. Optically addressable SLMs are particularly useful for applications like optical computing.

One method of controlling the operation of an SLM optically involves use of a thin film of photoconductive material. Figure 5-25 shows the operation of such a device fabricated using a sandwich consisting of ferroelectric ceramic material and a photoconductor between transparent electrodes. One such ferroelectric ceramic material is lanthanum-modified lead zirconate titanate, called PLZT. This material is transparent when it is polished in thin plates. Because of its ferroelectricity, it becomes electrically poled when voltage is applied to it. It also exhibits birefringence, which changes as the direction of electrical polarization changes. Thus, a structure consisting of PLZT between crossed polarizers will switch from transparent to opaque (or vice versa) in response to an applied voltage that suitably switches the direction of ferroelectric polarization.

In Figure 5-25, the ceramic element is originally poled with its polarization vector in the plane of the ceramic plate. The voltage drop across the ceramic depends

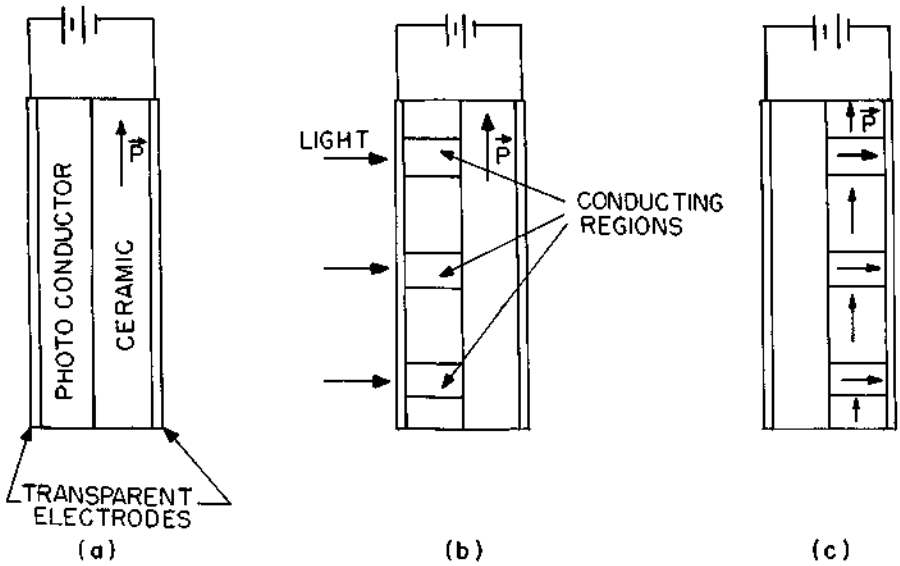


Figure 5-25 A ferroelectric-photoconductor switching element. In (a), the polarization vector **P** lies in the plane of the ferroelectric element. In (b), the photoconductor becomes transparent in regions where light is incident. In (c), the voltage has switched the direction of the polarization vector in those regions that the light addressed.

on the conductivity of the photoconductor. When no light is incident on the device, the photoconductor in not conductive, and most of the voltage drop appears across it. When light strikes the photoconductor in certain areas, it becomes conducting in those areas, and the voltage drop there falls to a low value. Then, the voltage drop appears across the ceramic in those regions where light intensity is high. This will cause the ceramic to switch its direction of electrical polarization, to a direction perpendicular to the plane of the plate. This change in electrical polarization is accompanied by a change in birefringence. If the material is between crossed polarizers, the device can switch from opaque to transparent in regions where the light intensity is high.

This example shows how an SLM can be optically controlled. The transmission of the SLM is controlled by the pattern of light intensity incident on it. In optical computing, such devices are useful for cascading or looping of devices.

Other types of materials can also be used with photoconductive materials to form optically addressable SLMs in a fashion similar to the example we have given. Electrooptic materials and liquid crystal light valves can be used in the same way.

P. Beam Homogenizers

A beam homogenizer, also called a beam integrator, is a device that smoothes out the irregularities in the spatial profile of a laser beam and converts the profile to a more uniform, flat-topped one. In most cases, a beam homogenizer is a multifaceted

mirror, with square facets. The different facets produce reflections that are directed at different angles so as to form a square beam with relatively constant power across the beam profile.

A variation intended for use with CO_2 lasers is a transmissive device. It consists of zinc selenide segments that break the beam up into a number of components and overlap the components at the target plane to produce a uniform intensity distribution.

Beam homogenizers are often used with excimer lasers, which frequently emit beams that have irregularities. They are also used with CO_2 lasers for applications like heat treating, for which one generally desires a flat beam profile.

Selected References

A. Mirrors

V. R. Costich, Multilayer Dielectric Coatings, in *Handbook of Lasers* (R. J. Pressley, ed.), Chemical Rubber Co., Cleveland, Ohio, 1971.

H. A. Macleod, *Thin-Film Optical Filters*, Macmillan, New York, 1986.

B. Optics

F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th ed., McGraw-Hill, New York, 1976.

W. J. Smith, *Modern Optical Engineering*, McGraw-Hill, New York, 1966.

C. Polarizers

M. Born and E. Wolf, *Principles of Optics*, Pergamon, Oxford, 1975, Chapter 14.

F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th ed., McGraw-Hill, New York, 1976.

D. Infrared Materials

P. W. Kruse, L. D. McGlauchlin, and R. B. McQuistan, *Elements of Infrared Technology*, Wiley, New York, 1962, Chapter 4.

S. Musikant, *Optical Materials: An Introduction to Selection and Application*, Marcel Dekker, New York, 1985.

E. D. Palik, ed., *Handbook of Optical Constants of Solids*, Academic Press, Orlando, FL, 1985.

W. L. Wolfe, Properties of Optical Materials, in *Handbook of Optics* (W. G. Driscoll and W. Vaughn, eds.), McGraw-Hill, New York, 1978.

E. Detectors

W. Budde, *Physical Detectors of Optical Radiation*, Academic Press, New York, 1983.

R. J. Keyes, ed., *Optical and Infrared Detectors*, Topics in Applied Physics, Vol. 19, Springer-Verlag, Berlin, 1980.

P. W. Kruse, L. D. McGlauchlin and R. B. McQuistan, *Elements of Infrared Technology*, Wiley, New York, 1962, Chapter 10.

K. J. Stahl, IR-Detectors: State-of-the-Art, Future Trends, *Photonics Spectra*, p. 95 (September 1989).

F. Modulators

I. C. Chang, Selection of Materials for Acousto-Optic Devices, *Opt. Eng.* **24**, 132 (1985).

C. C. Davis, *Lasers and Electro-Optics*, Cambridge University Press, New York, 1995, Chapter 19.

R. F. Enscoe and R. J. Kocka, Electro-Optic Modulation: Systems and Applications, *Lasers & Applications*, p. 91 (June 1984).
I. P. Kaminow, *An Introduction to Electrooptic Devices*, Academic Press, New York, 1974.
A. Korpel, *Acousto-Optics*, Marcel Dekker, New York, 1988.
Lasers & Optronics Staff, Crystalline Modulator and Q-Switch Specification Tables, *Lasers & Optronics*, p. 71 (November 1987).
R. P. Main, Fundamentals of Acousto-Optics, *Lasers & Applications*, p. 111 (June, 1984).
A. Yariv and P. Yeh, *Optical Waves in Crystals*, Wiley, New York, 1984, Chapters 8 and 10.

G. Light Beam Deflectors

I. C. Chang, Selection of Materials for Acousto-Optic Devices, *Opt. Eng.* **24**, 132 (1985).
M. Gottlieb, C. L. M. Ireland, and J. M. Ley, *Electro-Optic and Acousto-Optic Scanning and Deflection*, Marcel Dekker, New York, 1983.
A. Yariv, *Introduction to Optical Electronics*, Holt, Rinehart and Winston, New York, 1976, Chapters 9 and 12.
J. D. Zook, Light Beam Deflector Performance: A Comparative Analysis, *Appl. Opt.* **13**, 875 (1974).

H. *Q*-Switches

R. W. Hellwarth, Q Modulation of Lasers, in *Lasers, A Series of Advances*, Vol. I (A. K. Levine, ed.), Marcel Dekker, New York, 1966.
Lasers & Optronics Staff, Crystalline Modulator and Q-Switch Specification Tables, *Lasers & Optronics*, p. 71 (November 1987).

I. Nonlinear Optical Elements

R. W. Boyd, *Nonlinear Optics*, Academic Press, San Diego, 1991.
J. T. Lin and C. Chen, Choosing a Nonlinear Crystal, *Lasers & Optronics*, p. 59 (November 1987).
J. V. Moloney and A. C. Newell, *Nonlinear Optics*, Addison-Wesley, Reading, MA, 1991.
C. L. Tang *et al.*, NLO Materials Display Superior Performance, *Laser Focus World*, p. 87 (September 1990).
M. J. Weber, ed., *CRC Handbook of Laser Science and Technology, Volume III, Optical Materials, Part 1: Nonlinear Optical Properties/ Radiation Damage*, CRC Press, Boca Raton, FL, 1986.
F. Zernike and J. E. Midwinter, *Applied Nonlinear Optics*, Wiley, New York, 1973.

J. Optical Isolators

D. K. Wilson, Optical Isolators Adapt to Communication Needs, *Laser Focus World*, p. 175 (April 1991).

M. Beam Profilers

R. DeMeis, How to Select a Beam Profiler to Match Your Application, *Laser Focus World*, p. 133 (April 1995).

N. Optical Tables

J. R. Hobbs, Choosing the Right *Optical Table to Work with Your Application*, *Laser Focus World*, p. 103 (January 1995).

O. Spatial Light Modulators

B. Hill, Magneto-Optic Modulators, in Optical Computing, Digital and Sympbolic (R. Arrathoon, ed.), Marcel Dekker, New York, 1989.
D. R. Pape, Optically Addressed Membrane Spatial Light Modulators, *Opt. Eng.* **24**, 107 (1985).

Chapter 6 | Care and Maintenance of Lasers

High-power lasers can be self-destructive devices. Crystals, mirrors, and other optical components exposed to laser light with high irradiance all show degradation, more or less rapidly. This fact is of practical economic interest to users of lasers. Many of the components in high-power lasers are expensive, so that frequent replacement can be an economic hardship.

In addition to the components that can be damaged by the laser light, many other items undergo gradual (or sometimes catastrophic) degradation. These include items such as flashtubes, arc lamps, and electrodes carrying high current. Because the unique features of laser damage involve damage to nominally transparent materials by the laser light, much of the chapter will be devoted to that aspect of damage. Although many cessations of laser operation involve failure of the electrical power supply, there is nothing unique to laser applications in such failures, and we shall not emphasize this type of degradation.

Because of the economic importance of damage in lasers, we shall describe the damage processes and shall discuss some of the precautions and maintenance that should be employed in laser use.

A. Damage and Deterioration of Lasers

The generation of very high power pulses, often performed with solid state lasers like Nd:glass, can cause sudden, catastrophic damage to the laser materials themselves, and to the mirrors and other optical components in the beam. The damage phenomena are particularly important for Q-switched and mode-locked lasers, with short pulse durations. Lasers with longer pulse durations are less susceptible to damage.

Common lower-power gas lasers, like helium–neon and argon, are not vulnerable to dramatic catastrophic damage of the type discussed in this section. Such lasers undergo other types of deterioration, which we shall discuss later.

1. DAMAGE IN OPTICAL MATERIALS

High-power laser light can produce damage in materials that are nominally transparent to the light at low intensity. The initiation of optical damage or optical breakdown occurs at some threshold value of laser irradiance. At values below the threshold, the light is transmitted, apparently without effect on the material. When the irradiance is increased above the threshold, catastrophic breakdown occurs, damaging the material. The damage can take the form of pitting, cracking, and vaporization of material. It is often accompanied by a flash of light and by a sharp sound. Thereafter, the optical component, whether mirror, laser rod, window, or lens, has its performance as an optical element much degraded. The damage grows worse if additional laser pulses strike the component.

As an example of the catastrophic effects of optical breakdown, Figure 6-1 shows a photograph of a potassium chloride surface damaged by a high-power CO_2 laser pulse delivering an irradiance above 10^8 W/cm^2. At lower values of irradiance, potassium chloride is transparent to CO_2 laser light, and no damaging effects occur. As one increases the irradiance, one reaches a threshold where a bright spark may be observed. Cracking and material removal such as illustrated in the figure occur. This type of damage is uniquely laser related because other optical sources cannot deliver irradiance exceeding the damage threshold.

The value of the breakdown threshold depends in a complicated fashion on many factors, including the nature of the material, the wavelength of the laser light, the duration of the exposure, the previous history of the sample, the presence of impurities and imperfections in the sample, and the state of surface finish. The value of the breakdown threshold also shows statistical fluctuations. The spread of the statistical fluctuations often is narrow enough that one may define the breakdown threshold with reasonable precision.

There is voluminous literature describing laser-induced damage and the physical mechanisms that produce it. We will not provide a complete review of the phenomena involved; rather, we shall summarize some of the pertinent conclusions.

Early work on identifying damage mechanisms was hindered by several problems. These included uncertainties about temporal substructure in the laser pulse, the presence of absorbing particles in the samples, and a phenomenon called self-focusing. In self-focusing, the electric field of the light wave causes a change in the index of refraction of the material, in such a way that the light is constrained to travel in a small filament through the material. The presence of self-focusing means that the laser irradiance is much increased above the value expected without self-focusing. Many of the early measurements of laser damage threshold were probably actually measurements of the threshold for self-focusing.

In many other cases, breakdown is initiated by the presence of small absorbing inclusions in the samples. Preparation of material of high perfection and the use of focused laser beams, both reducing the probability of striking an imperfection, have led to measurements of the intrinsic breakdown threshold for many materials. The intrinsic threshold is the threshold characteristic of the material itself, without self-focusing or the influence of absorbing inclusions or other defects.

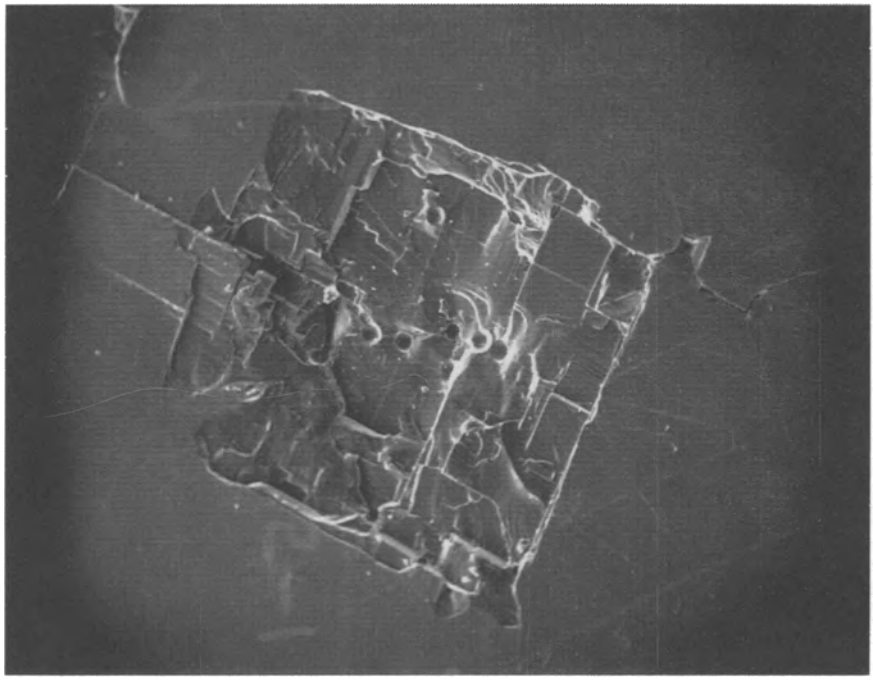


Figure 6-1 Scanning electron microscope photograph of damage at the surface of a KCl sample. The damage was produced by a CO_2 laser pulse delivering about 10^8 W/cm^2 to the surface. At lower irradiance, the KCl is transparent, and no damage occurs. The width of the area shown is 1 mm.

It is now generally believed that optical breakdown (at least intrinsic breakdown) is similar to dielectric breakdown caused by a very high dc electric field. The mechanism for dielectric breakdown of an insulating material includes avalanche or cascade ionization, a process in which electrons are accelerated in the electric field of the light wave, gain energy, collide with bound electrons, and excite them across the energy gap of the material. These electrons then also may gain energy from the light wave and cause further excitation. This process starts with a very small number of free electrons in the material. Cascade multiplication yields an absorbing plasma into which a significant fraction of the laser energy can be coupled, and eventually this produces damage in the material. At low frequency, the ability of an electron to gain sufficient energy to cause excitation of additional free electrons is limited by electron–phonon collisions. Unless the electric field is very high, the electrons will lose energy in these collisions before they can excite other electrons. When it is the field of a light wave, the electric field must be even higher, so that the electron can gain enough energy within one-half cycle of the oscillating field. Otherwise, the field will change direction and the electron will be decelerated before it can cause further excitation of free electrons.

Studies have shown that the breakdown values of electric field for alkali halides are essentially the same at dc, 10.6 μm and 1.06 μm [1]. The breakdown threshold expressed in terms of the electric field in the light wave may be related to the breakdown threshold for dc dielectric breakdown by the relation

$$E_t(\omega) = E_t(0)(1 + \omega^2\tau^2)^{1/2} \tag{6.1}$$

where $E_t(\omega)$ is the electric field needed to produce breakdown at frequency ω, $E_t(0)$ is the dc breakdown field, and τ is an effective electron collision time, which may be approximately 10^{-15} sec. When the frequency of the light wave becomes higher than $1/\tau$, the threshold will increase. For alkali halides, this should occur when the wavelength becomes shorter than 1 μm.

Multiphoton absorption of laser energy is also implicated in breakdown of nominally transparent optical materials. Models that include both avalanche ionization and excitation of electrons across the bandgap of an insulating material by absorption of several quanta of photon energy give better agreement with experiment than models that include only one of these mechanisms [2].

When the photon energy becomes greater than half the bandgap energy, two-photon absorption becomes possible. This is an effective absorption mechanism when the irradiance is very high, so that the threshold for damage decreases.

Another observation has been that the threshold for damage usually is lower at the surface of a material than in the bulk. This has been related to the presence of small imperfections, such as scratches, pits, and voids, near the surface [3]. The discontinuity in dielectric constant at a surface can enhance the electric field. The enhancement depends on the geometry of the defects present near the surface. The enhancement of the electric field leads to a lower apparent breakdown threshold. Suitable polished surfaces free of defects have shown breakdown thresholds comparable to those of bulk material. As a rough rule of thumb, the damage threshold for a surface varies as the inverse square root of the root mean square of the surface roughness.

In addition to damage, there are problems associated with absorption of laser energy by lens and window materials with resultant deformation of the material and degradation of the quality of the beam that is transmitted. Such effects that degrade system performance may be encountered at values of irradiance well below those required for catastrophic damage. The problem is especially severe for infrared lasers, for which the available materials may have higher values of absorption than materials used with shorter-wavelength lasers. A spatially inhomogeneous laser beam causes a temperature gradient, which produces nonuniform changes in the thickness and index of refraction of the material. The resultant increased physical thickness near the center of a window material causes it to become a positive lens having aberrations, birefringence, and a finite focal length. The effect of the temperature-dependent change in index of refraction varies with the material. Covalent materials generally yield a positive contribution and make the effect worse. Ionic crystals like NaCl have a negative effect, which at least partially offsets the contribution of the increased physical thickness.

This discussion will serve as a brief introduction to the complex range of phenomena encountered in catastrophic breakdown of nominally transparent materials by high laser power. It is of economic importance because lenses, windows, and other components placed in the path of a laser beam will be damaged if the irradiance becomes too high, and they will have to be replaced.

We next consider the degradation of components used in lasers.

2. *SOLID STATE LASER MATERIALS*

Solid state lasers like Nd:glass are often used for generation of short, high-power pulses. It is this application that generates the most serious problems of laser damage. The laser rods in solid state lasers can be damaged by the high light intensity. The damage often takes the form of voids within the rod or pits on the ends of the rod. Once damaged, the output power from the laser decreases or the laser may fail to operate. The rod, one of the more costly elements in the laser, must then be replaced.

The damage again has a threshold, below which the effects are small. Figure 6-2 shows data on the damage threshold for a particular type of Nd:glass that has been used in high-power lasers. This represents a glass with relatively high damage threshold. Once damage has been produced, the output of the laser rod is much decreased. The figure expresses damage threshold in terms of laser fluence as a function of pulse duration. The threshold in terms of fluence increases as the pulse duration increases. The numbers on the curve indicate the laser irradiance at the particular points. The irradiance that can be tolerated without damage decreases as the pulse duration becomes longer.

The data presented in Figure 6-2 indicate that if a given laser fluence is required, one should operate at relatively low values of peak power and make the pulse duration as long as possible. But if a high value of irradiance is needed, the pulse duration should be shortened in order to reduce the total fluence.

Damage may occur both internally within the rod and at the ends of the rod. Usually, damage begins near the output end of the rod and progressively grows inward toward the opposite end with further pulsing of the laser. The damage threshold for a given laser material shows a great deal of scatter because of variation among individual specimens.

The data shown in Figure 6-2 represent damage produced within a single laser pulse. It is also possible to produce damage from a number of shots, even if one remains below the threshold for damage in a single shot. Table 6-1 gives some data on

Table 6-1 **Number of Shots before Damage (30 Percent Decrease in Output) Occurs in Ruby Rods, for 10-Nanosecond-Duration Laser Pulses** [a]

Irradiance (W/cm2)	3×10^9	10^9	3×10^8	10^8	3×10^7
Number of shots	1	25	500	5000	100,000

[a] From F. P. Burns, *IEEE Spectrum*, p. 115 (March 1967).

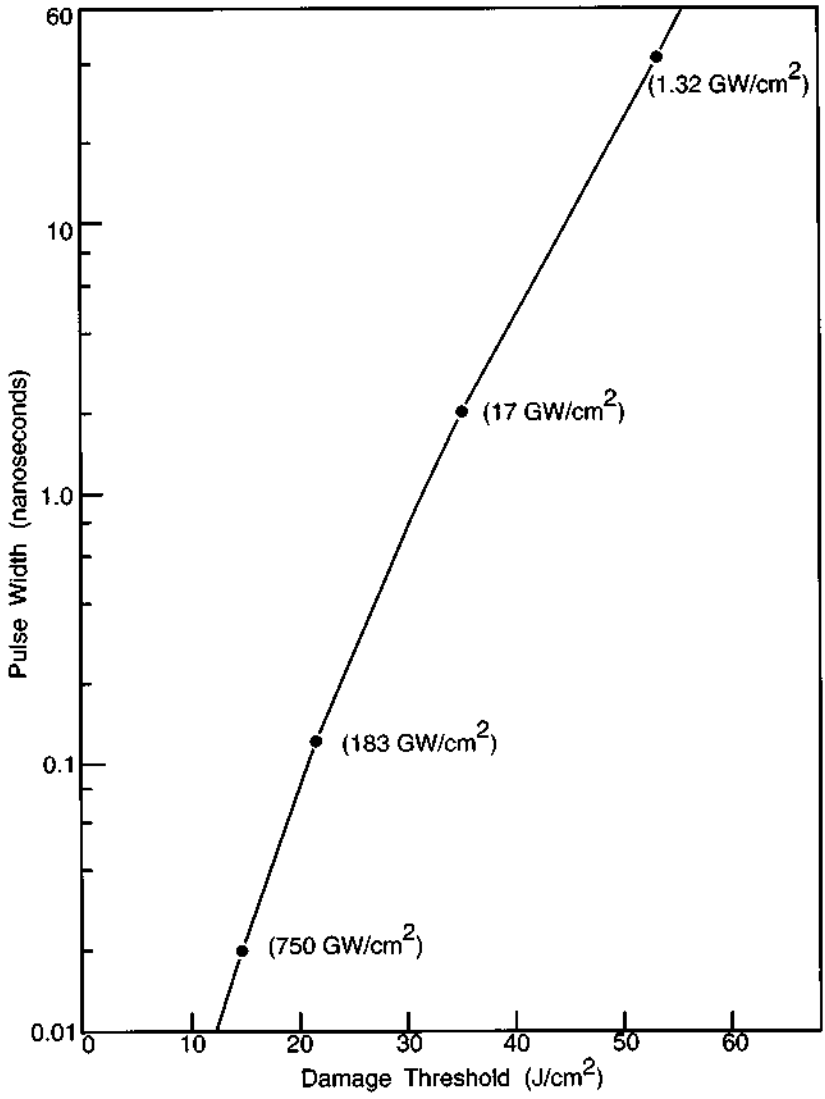


Figure 6-2 Threshold values for optically induced damage in ED-2 Nd:glass as a function of laser pulse duration. The numbers on the curve give the irradiance at the particular points. (Based on data from Owens-Illinois, Inc.)

the number of shots before damage can be expected in a ruby rod, as a function of irradiance. The data are appropriate for a pulse duration of 10 nsec. The data represent the number of pulses that could be extracted from a ruby laser under constant excitation conditions before the output was reduced by 30 percent. At an irradiance around 3×10^9 W/cm^2, damage is produced within a single pulse. As the irradiance

is lowered, the number of pulses that can be extracted without damage increases considerably.

Figure 6-3 shows a photograph of internal damage within a ruby rod, produced by about 20 shots at a power around 200 MW. Internal damage such as this commonly takes the form of bubbles or voids, which are made visible by scattering of He–Ne laser light. In addition, pitting and cracking of the surface are common manifestations of damage.

Catastrophic damage of the type discussed here is especially important for lasers with very high peak power and short pulse duration. An example involves the program on laser-assisted thermonuclear fusion, in which extensive efforts to reduce damaging effects have been employed. These phenomena are less often encountered now with commercial solid state lasers, for which the manufacturers now ensure that components are not exposed to laser irradiance above the damage threshold.

3. *MIRRORS AND WINDOWS*

Damage to laser mirrors is crucial. In any high-power laser, the mirrors undergo degradation, more or less rapidly. The operating budget for such a laser must include an allowance for periodic mirror replacement.

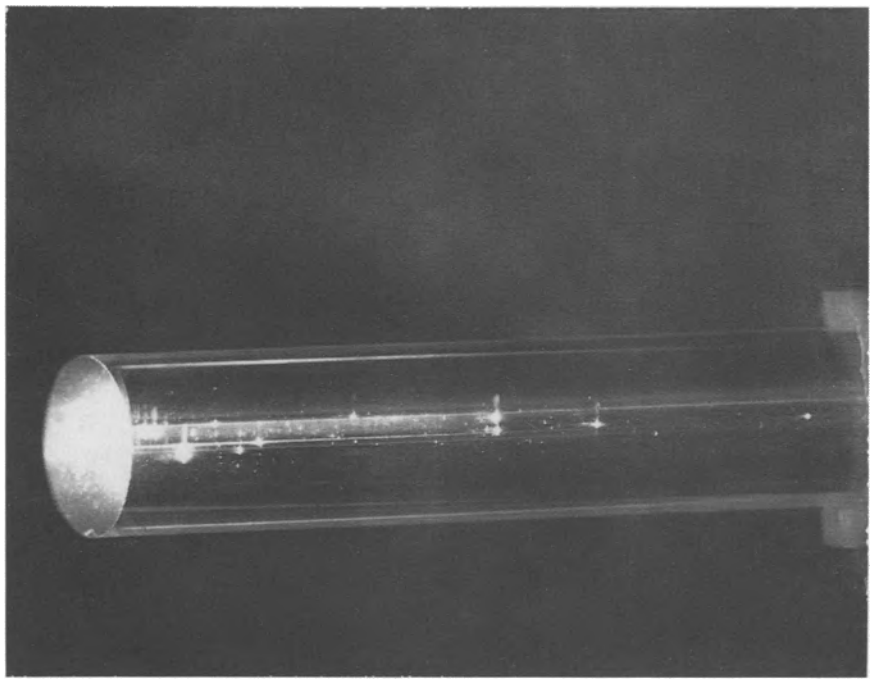


Figure 6-3 Damage in a ⅝-in.-diameter ruby rod occurring after 20 shots at 200 MW power. The damage is viewed by scattering of He–Ne laser light traversing the rod axis. (From J. F. Ready, *Effects of High-Power Laser Radiation*, Academic Press, New York, (1971).)

For the earliest lasers, the mirrors were metallic coatings, often vacuum-deposited or chemically deposited silver. These mirrors degraded very rapidly. In addition, such mirrors often had unacceptably high values of loss because of absorption and scattering.

Modern lasers use multilayer dielectric coatings, as have been described in Chapter 5. Such coatings are fabricated by successive evaporation of a number of layers of dielectric material, each layer with an optical thickness of one-quarter wavelength. Two different materials, one with high index of refraction and one with low index, are used in alternating layers.

A large amount of effort has been devoted to developing improved coating technology over the decades since the invention of the laser, and substantial progress has been made in developing high-quality laser mirrors with increased damage thresholds. Despite this, the mirrors still remain one of the most vulnerable and most frequently damaged components in lasers. Figure 6-4 shows a typical example of laser-induced damage in a mirror coating, in which the layers of the coating are detached and separated from the substrate.

Damage thresholds for vacuum-deposited thin-film materials commonly used in multilayer mirrors for use in the near infrared, visible, and ultraviolet portions of the

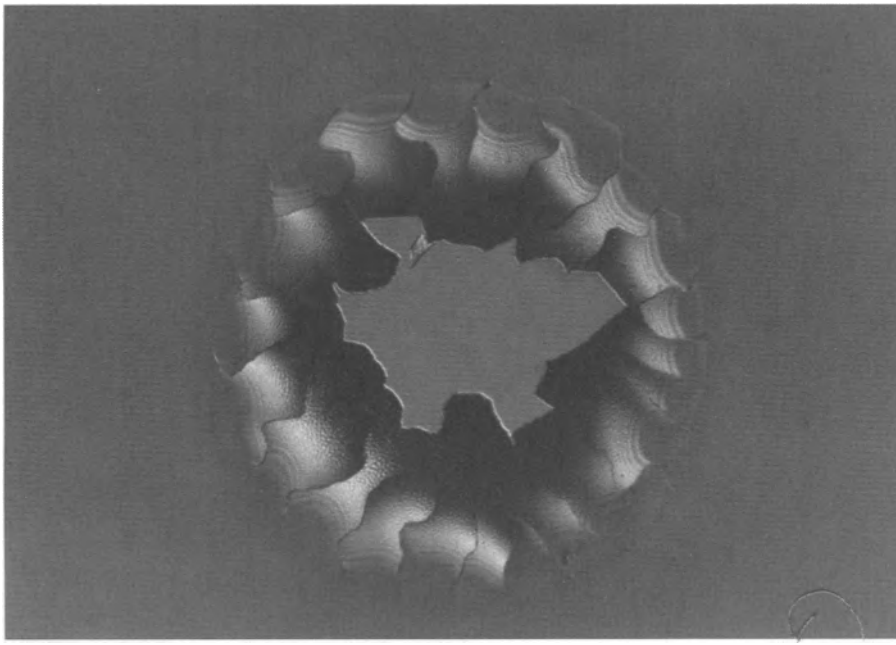


Figure 6-4 Laser-induced damage to a multilayer dielectric mirror coating, with the layers of the coating separated from the mirror substrate. (Photograph courtesy of Montana Laser Optics, Inc.)

spectrum are given in Table 6-2. These films were deposited on silica substrates in a cryogenic vacuum system using good commercial practice. The damage thresholds were measured as a function of wavelength, pulse duration, and coating thickness. The results indicate that the damage threshold (in terms of fluence) increases with increasing pulse duration, increases with increasing wavelength, and decreases with increasing thickness. Systematic variations between materials are also apparent. The important feature is that the damage thresholds are commonly as high as several J/cm^2, and sometimes as high as several tens of J/cm^2, for nanosecond duration pulses. This represents a substantial improvement over the coatings available near the start of the laser era. Multilayer mirrors made from these materials will probably have damage thresholds intermediate between the thresholds of the two materials that make up the mirrors. Multilayer dielectric coatings with damage thresholds well in excess of 10 J/cm^2 at 1.06 μm have in fact been demonstrated [4].

Multilayer dielectric mirrors for use with CO_2 lasers must employ materials that are transparent at the long-infrared wavelength of that laser. These mirrors are

Table 6-2 **Damage Thresholds (J/cm^2) for Thin-Film Optical Coatings**[a]

Material	Thickness[b]	Wavelength and pulse duration							
		1.06 mm		0.53 mm		0.353 mm		0.26 mm	
		5 nsec	15 nsec	5 nsec	15 nsec	5 nsec	15 nsec	5 nsec	15 nsec
MgF_2	1	9.2	13.1	3.6	7.9	3.6	4.2	2.9	2.4
	½	—	15.8	6.9	12.1	—	—	—	2.0
	¼	11.5	16.9	7.2	14.4	2.7	4.7	1.9	2.9
	⅛	27.4	37.0	8.9	13.9	2.3	5.4	1.3	3.0
Al_2O_3	1	11.4	15.1	—	13.2	3.2	6.6	1.5	2.3
	½	—	16.3	6.5	11.7	—	—	—	—
	¼	13.5	20.1	7.3	13.2	2.7	6.0	2.2	2.0
	⅛	20.0	21.5	8.5	14.0	2.9	5.9	1.7	2.3
MgO	1	10.2	9.8	—	11.4	—	—	—	—
	½	14.9	15.0	6.1	10.6	—	—	—	—
	¼	—	—	6.8	16.2	2.8	3.9	1.3	2.3
	⅛	12.2	12.8	10.6	14.2	2.8	4.3	1.6	2.6
TiO_2	1	8.8	—	4.3	—	0.1	—	—	—
	½	11.4	—	4.8	—	—	—	—	—
	¼	14.9	13.8	6.5	11.8	0.1	—	—	—
	⅛	10.0	9.8	6.4	10.8	0.6	—	—	—
SiO_2	1	—	—	—	—	—	—	—	—
	½	—	7.8	—	—	—	—	—	—
	¼	22.4	45.0	9.4	19.6	5.6	10.7	0.9	1.6
	⅛	41.4	47.3	21.1	48.4	—	13.5	2.2	2.5

[a]From T. W. Walker, A. H. Guenther, and P. E. Nielsen, *IEEE J. Quantum Electron.* **QE-17**, 2041 (1981).
[b]In units of 1.06 μm.

fabricated using materials like CdTe, ZnTe, and ThF_4. They are subject to damage, which may appear as pits on the coating surface after many laser pulses, even when no damage was observable after a single shot. Figure 6-5 shows a photograph of a coated germanium mirror for a CO_2 TEA laser, after several hundred thousand laser pulses at an irradiance around 2×10^6 W/cm^2. The laser output had decreased about 30 percent with the mirror in this condition.

The high-reflectivity mirrors for CO_2 lasers often use metallic mirrors, such as coated copper and molybdenum. This use is feasible because the reflectivity of metals becomes very high at the CO_2 laser wavelength. Such mirrors exhibit reasonably good damage characteristics.

4. GAS LASERS

For relatively low power gas lasers, lifetimes of many thousands of hours are common. The eventual failure arises from gas depletion or leaks, rather than from damage produced by the laser beam. Helium–neon lasers in particular often eventually fail because of depletion of the helium by diffusion of the small helium atoms through the walls of the tube.

Argon lasers require very high current density. The argon ions may be driven into the walls of the tube and become trapped, with the result that the argon gas pressure drops below the optimum value for laser operation. Commercial argon lasers incor-

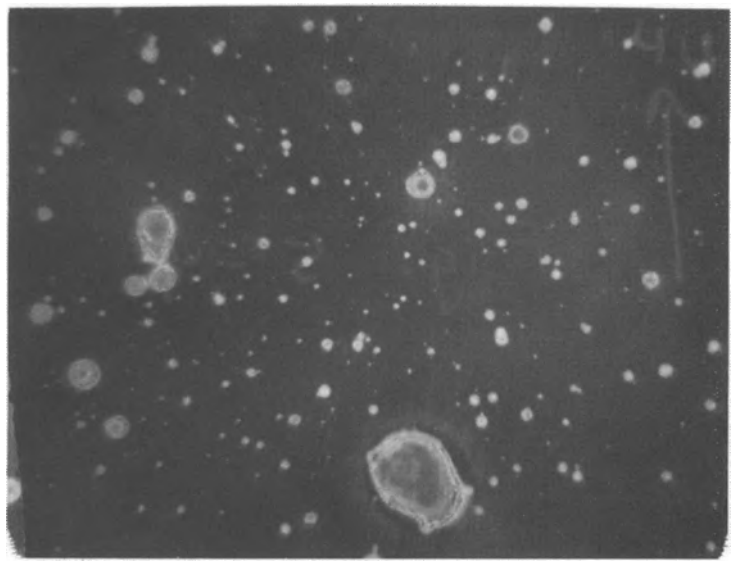


Figure 6-5 Damage appearing on a coated germanium output mirror for a CO_2 TEA laser. The damage developed after several hundred thousand pulses at 10 MW output over a 1 in. area and reduced the laser power by about ⅓. The width of the area shown in the photograph is 3.8 mm.

porate an argon reservoir and provision for repressurizing the tube in order to counteract the argon depletion.

Sealed CO_2 lasers may suffer from dissociation of the carbon dioxide in the electrical discharge and resultant buildup of impurity gases like carbon monoxide. As a result, the output power of the laser can drop after a few hundred hours of operation. Most sealed CO_2 lasers now incorporate catalysts or other materials that provide a reaction path to allow the decomposition products to react and regenerate the carbon dioxide. With this provision, sealed CO_2 lasers now operate for many thousands of hours without degradation of the output.

Higher-power CO_2 lasers with flowing gas avoid the problem of gas dissociation but are subject to degradation mechanisms such as erosion of electrodes and mirror damage. With a suitable maintenance program to identify and replace deteriorating components, a CO_2 laser can have essentially unlimited lifetime, with minimal downtime.

5. *SEMICONDUCTOR LASERS*

Semiconductor lasers are subject to damage from a number of causes. Possibly the most serious cause involves catastrophic optical damage (COD), which occurs when the laser power at the surface of the laser diode becomes too high, greater than about 10^6 W/cm^2. Absorption of laser light at the surface leads to heating, which in turn increases the absorption further. The result is a feedback process that leads to high temperature and melting of the diode surface and destruction of the laser.

When COD occurs, the output power of a laser diode as a function of drive current may appear as in Figure 6-6. The output increases with increasing current above the threshold current. When the output becomes too high, damage occurs and the output drops abruptly. The surface damage may prevent laser operation completely, or it may allow some operation at reduced power.

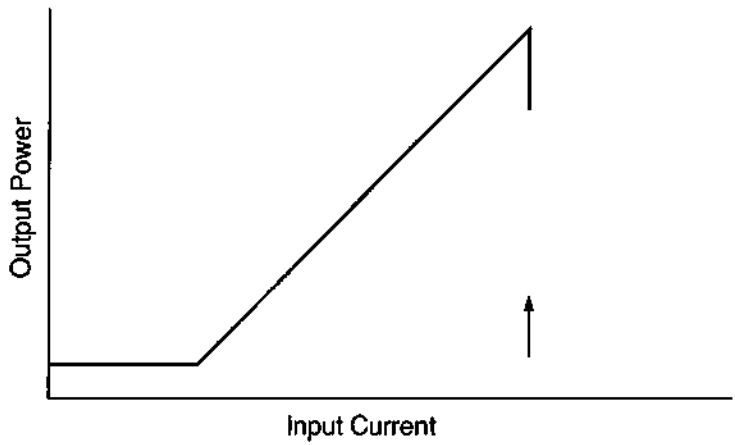


Figure 6-6 The output of a laser diode as a function of drive current. The threshold for catastrophic optical damage is indicated by the arrow.

Even without COD, the output of a semiconductor laser gradually decreases after prolonged operation, probably due to buildup of damage around faults or defects originally present in the material. This gradual degradation can be minimized if the current is kept reasonably small and the operating temperature is kept low. With these precautions, the lifetime for a semiconductor laser may be quite long, years or even decades. Sometimes manufacturers have claimed, on the basis of accelerated aging tests, that lifetimes could exceed 100 years.

The gradual degradation is accelerated by increasing temperature. The operating lifetime of a laser diode is significantly reduced by operation at elevated temperature. The lifetime is reduced by a factor that varies with absolute temperature T as $\exp(E_a/kT)$, where E_a is an activation energy, typically around 0.5 to 0.7 eV, and k is Boltzmann's constant. According to this dependence, an increase in operating temperature of 40 K will decrease the lifetime by a factor of around 30.

Another hazard to semiconductor diode lasers is excessive voltage, which may arise from power supply transients. Many power supplies produce voltage spikes during turn-on or turn-off. These electrical transients may be large enough to cause damage to the semiconductor laser.

Similarly, static electrical discharges are potentially damaging to semiconductor lasers. Electrostatic discharges are a major cause of unexpected semiconductor laser failure. In practice many semiconductor lasers fail because of "electrocution" in some form.

The probability of diode laser destruction by electrostatic discharge or other electrical transients depends on the skill and experience of the user, and on the particular application. If a semiconductor laser is used under carefully controlled conditions of temperature and current, it may well last for a very long time. But in the hands of an inexperienced operator, the lifetime of a semiconductor laser may be not years, but a few seconds.

6. DYE LASERS

Dye lasers suffer degradation because of breakup of dye molecules by the pump radiation. The large dye molecules are subject to photodissociation by the pump laser light. This leads to decrease of the dye laser output over time. The dye must be replaced with fresh material in order to restore the output of the laser. The shorter the wavelength of the pump laser radiation, the worse the problem becomes. For typical pump lasers with wavelength longer than 300 nm, the dye lifetime (specified as 50 percent reduction in output) may be in the range 10 to 50 watt-hours of pump laser radiation per liter of dye solution. Thus, if one has 1 L of dye excited by 2 W of pump laser power, the dye may last between 5 and 25 hours. For pump laser wavelengths shorter than 300 nm, the dye lifetime may be shorter.

7. DETECTORS

Optical detectors are commonly employed as components in laser systems. Detectors differ in their application from many other components because the detectors

are designed to absorb much of the energy incident on them. Many detectors have large values of absorption coefficient, often in excess of 10^4 cm^{-1}, so that the incident energy is absorbed in a small volume at the surface of the detector. Additionally, detectors are predominantly fabricated using semiconductor materials that have lower melting temperatures than many other optical materials. Thus, optical detectors are particularly vulnerable to damage by the laser beam. A large amount of work has been performed to investigate the damage of detector materials. This includes experimental determination of damage thresholds and mathematical modeling of the damage processes.

Laser damage in detector materials is usually irreversible damage in the form of surface melting or vaporization. In some cases, the electrical leads to the detector may become separated. A damaged detector may cease to function entirely, or it may continue to operate with lowered responsivity.

The results of extensive testing to determine the damage thresholds of a variety of detector materials is summarized in Figure 6-7. The detector materials include those commonly employed in the visible, near infrared, and far infrared portions of the spectrum, for a range of exposure times ranging from 10^{-7} to 1 sec.

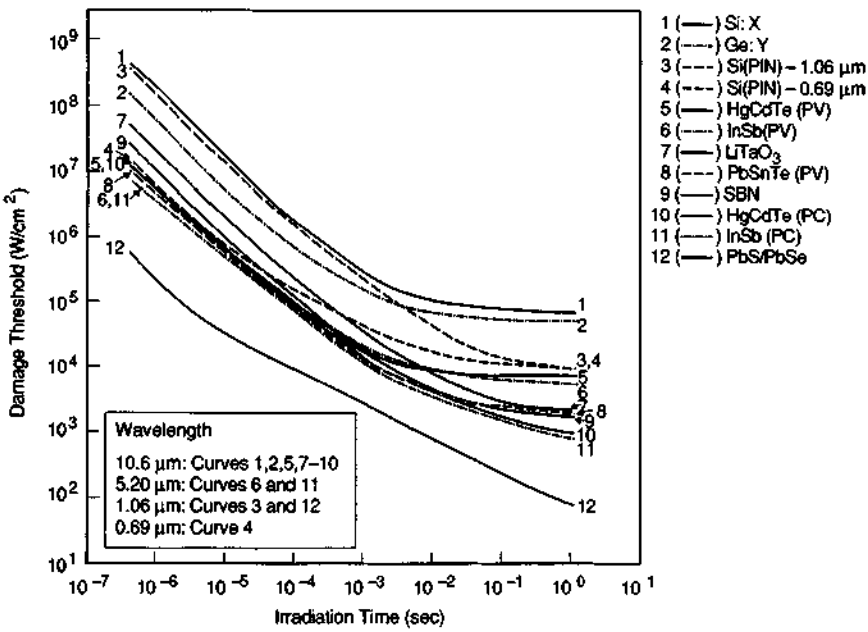


Figure 6-7 Experimental damage thresholds for optical detector materials, as a function of exposure time, for a variety of detector materials, as indicated. The wavelengths appropriate to each curve are identified, and these include wavelengths ranging from the visible to the far infrared portions of the spectrum. The notation Si:X and Ge:Y refers to doping of the host material with dopants such as Ga and As. (From F. Bartoli *et al.*, *Appl. Opt.* **16**, 2934 (1977).)

Near the left of the figure, the damage threshold, in terms of irradiance, decreases with increasing duration. In this region of short pulses, there is little time for thermal conduction to carry heat away from the region in which it is absorbed, and the thresholds are dominated by the amount of energy absorbed per unit volume. The damage threshold is the irradiance required to deliver enough energy to raise the material to its melting temperature. Toward the right of the figure, the curves level off and become independent of pulse duration. In this region, the irradiance is balanced by thermal conduction out of the absorbing region. The threshold is determined in this region by the power per unit area rather than by the energy per unit area.

The figure indicates that under easily attainable levels of laser irradiance, common optical detector materials may be permanently damaged.

8. *FLASHTUBES AND ARC LAMPS*

The flashtubes and arc lamps that are used for optical pumping of high-power solid state lasers also have finite lifetime. The lifetime is reduced as the input energy is increased. At some value of input energy, a flashtube used for pumping pulsed lasers will explode in a single pulse. The explosion energy depends on the diameter and length of the tube. Figure 6-8 shows the explosion energy for flashtubes as a function of the diameter of the tube and the duration of the discharge of electrical energy through the tube. The pulse length may be adjusted by varying the circuit parameters, including capacitance and voltage of the power supply and the value of the circuit inductance. The explosion energy is expressed in terms of the energy input to

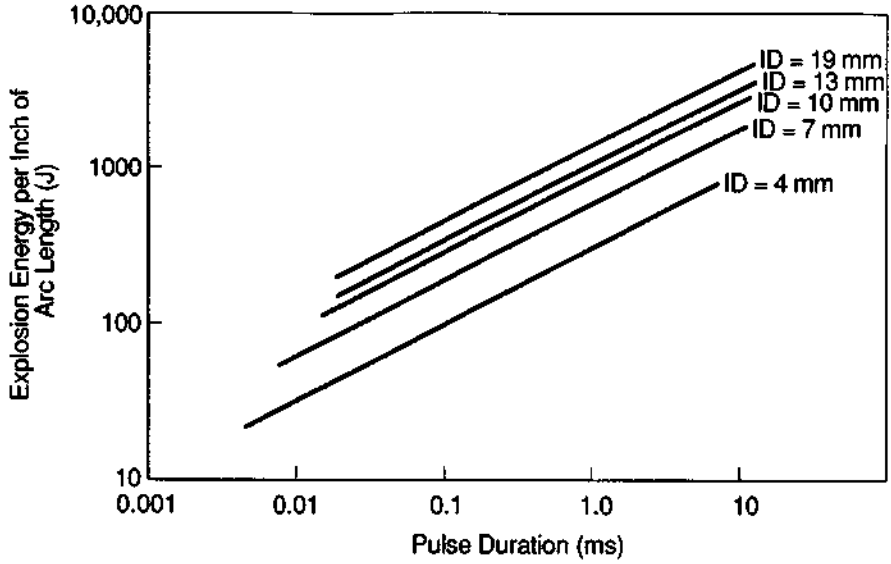


Figure 6-8 Flashtube explosion energy as a function of pulse duration, expressed as the duration at one-third peak amplitude of the electrical input to the tube. The tube inside diameter (ID) is a parameter. (From G. A. Hardway, *MicroWaves* **5**, 46 (1966).)

the tube per unit length of the tube. This value will cause the flashtube to explode in one pulse.

To obtain long life for flashtubes, the tube must be operated at values of the electrical input energy much below the explosion energy. Figure 6-9 shows how the lifetime is increased as the electrical loading is reduced. To obtain reasonably long lifetime, one should operate below 50 percent of the explosion energy. When the electrical input is restricted to lower values, perhaps 10 to 20 percent of the explosion energy, the lifetime may rise to millions of pulses. However, the user's desire to obtain high output energy from the laser often requires larger inputs to the flashtube. Some trade-off must generally be made between lifetime of the flashtube and the output of the laser.

In the regime where the lamp is operated well below the explosion energy, the failure mechanism is generally not an abrupt failure; rather, it is a gradual decrease in light output because of darkening of the interior surfaces of the tube as material is evaporated from the electrodes and redeposited on the flashtube walls.

The arc lamps used for excitation of continuous solid state lasers are also subject to degradation, primarily because of evaporation and sputtering from the metal electrodes, with resultant deposition of material on the interior surfaces of the lamp. This darkens the surfaces and reduces the light output. As might be expected, the useful lifetime shortens as the input to the lamp increases. Figure 6-10 shows how lamp lifetime, defined as the time during which the lamp output exceeds 90 percent

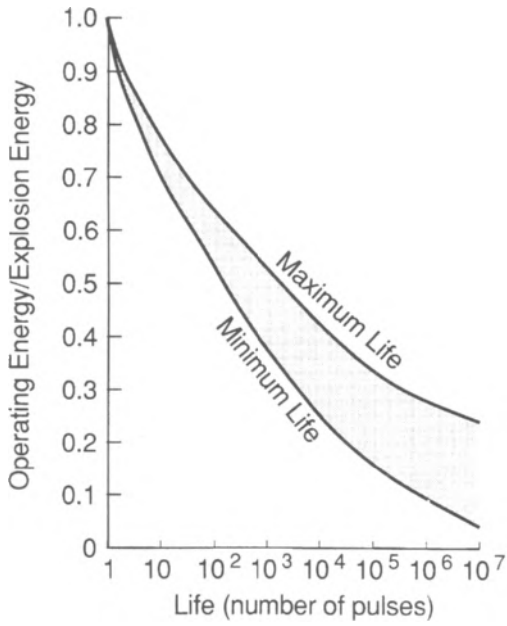


Figure 6-9 Flashtube life as a function of input energy. (From ILC Technical Bulletin 1, *An Introduction to Flash Tubes*, ILC Technology, Sunnyvale, CA.)

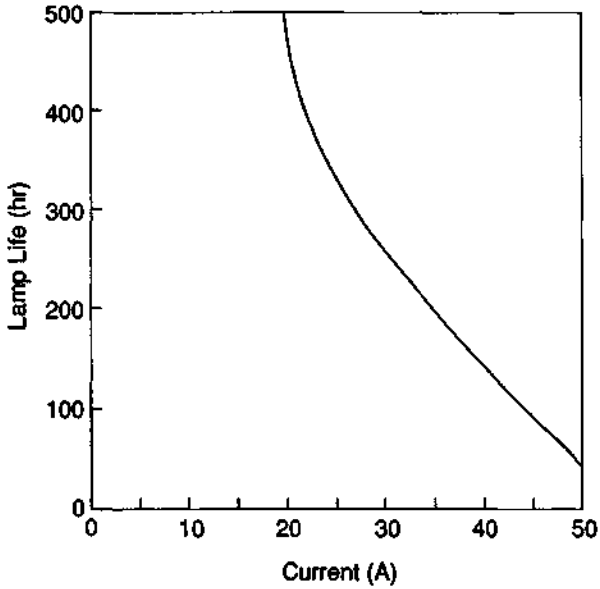


Figure 6-10 Lifetime versus input current for a commercial krypton arc lamp. (Data from Quantronix Corp.)

of the rated value, decreases as the current through the lamp increases. The figure is relevant to a krypton arc lamp used to pump Nd:YAG lasers.

9. *SUMMARY*

At high levels of power, lasers become self-destructive. Some damage must be expected. The user of a high-power laser must have a suitable operating budget for maintenance and replacement of damaged components. Probably the most serious catastrophic damage would be to a laser rod, because the cost of the other elements is usually lower. Components like mirrors may be replaced at lower expense, although if replacement is frequent, this can have a serious economic effect.

In applications requiring high peak power and short pulse duration, the lifetime of the components should be estimated beforehand in order to determine probable operating costs. The data presented here indicate probable lifetimes for certain operating conditions. In general, the lifetime may be extended by reducing the irradiance. Thus, the user must often make a trade-off between operating cost and high irradiance.

B. Care and Maintenance

In many lasers with relatively low output power, like helium–neon lasers, the tubes are sealed, and there is little that the laser user can do to extend its life beyond what the manufacturer has already done.

For commercial high-power lasers, the manufacturer usually limits some of the parameters that can induce damage. Thus, most commercial lasers have maximum values for the input current, input energy, and so forth. These maximum values are set at levels that should eliminate rapid catastrophic damage and that should allow an operating life that meets the manufacturer's warranty.

The main items of interest in high-power solid state lasers are the rods of glass or YAG, the mirrors, and the flashtubes or arc lamps. We have already discussed the mechanisms of damage in these components and presented data on damage thresholds. In this section, we discuss some methods for minimizing the damage and for keeping lasers operating at their optimum output.

1. LASER RODS

The most important consideration is to keep the laser irradiance below the damage threshold. It must be recognized that damage thresholds are usually quoted in terms of average irradiance over the spatial cross section of the beam and over the temporal duration of the pulse. In many cases, instantaneous values may be higher. This is particularly true if mode-locking occurs. The peak value of power may be much higher than the value averaged over the pulse. Thus, unless one specifically desires mode-locked behavior, one should suppress mode-locking in order to reduce the probability of damage.

Also, one should ensure that the spatial profile of the beam is relatively uniform and does not contain hot spots where the irradiance becomes very high. Bleachable dye Q-switches are troublesome because they tend to bleach unevenly and produce irregular beam profiles. The use of electrooptic or acoustooptic Q-switches leads to beams with smoother spatial profiles.

One common procedure that helps avoid damage in high-power lasers is the use of a master oscillator, power amplifier (MOPA) configuration. The original laser pulse is generated in a master oscillator with small cross section. The beam, with relatively low power, can have high quality with a smooth spatial profile. The beam is then allowed to pass through several stages of amplification. Each successive stage has progressively larger cross section, so that the irradiance never becomes too large. The spatial quality of the beam is preserved in these amplifier stages, which are essentially lasers without mirrors. The laser material and optical pumping are present, but there is no feedback that would allow the amplifier stages to operate as independent lasers. As the pulse passes through the amplifiers, the energy stored in the amplifiers is added to the beam from the oscillator by the stimulated emission process. The amplifiers may employ laser rods with diameter greater than that of the oscillator. As the beam cross section becomes larger, the laser material may then take the form of slabs or disks, in order to keep increasing their area. Between stages, the beam may be expanded by lenses to fill the aperture of the next stage, always keeping the irradiance below the damage threshold.

It is essential to keep light from being reflected backward in a MOPA configuration. Any reflected light traveling backward will be amplified as it goes through

stages with progressively smaller area. One could quickly exceed the threshold for damage. Thus, in a MOPA device, optical isolators are employed between stages to ensure one-way beam propagation.

2. *MIRRORS*

Mirrors are particularly susceptible to damage by high-power lasers. The multilayer dielectric mirrors that are commonly used will degrade in some hundreds or thousands of pulses at high power levels. Mirror surfaces must be kept scrupulously clean and free of dust. Any dust particles present on the mirror surface (or any other surface in the laser) will be vaporized and will serve as a nucleation point for damage to begin. Fingerprints on mirror surfaces must also be avoided. As an example of the effect of dust, a CO_2 laser mirror coating survived irradiation by 1000 W/cm^2 for 20 sec when it was clean. When it was purposely covered with dust, it damaged below 100 W/cm^2 [5].

Cleaning of mirror surfaces is important to prolong their life. Before we discuss cleaning, we should comment on the different types of mirror coatings. Multilayer dielectric coatings fabricated by vacuum deposition may be considered to be hard, semihard, or soft. Hard coatings contain refractory materials such as silicon dioxide and titanium dioxide. These materials are durable, resist scratching, and are not soluble in water. The mirrors used in modern ultraviolet, visible, and near infrared lasers are generally of this type. Semihard coatings use materials such as thorium fluoride and zinc selenide, which are often used in the long-wavelength infrared. These materials are less hard and durable, but they provide low infrared absorption. They are less resistant to scratching and may have some water solubility. Soft coatings are not used in modern lasers, but some mirrors produced in the early history of lasers were of the soft type. Such mirrors were not resistant to scratching and could be water soluble.

Hard coatings may be cleaned by the following procedure: They may be washed gently in warm soapy water with a mild liquid detergent, if there are large amounts of contamination on the surface. Such cleaning may not always be necessary. The final cleaning step (whether or not washing was needed) involves use of a solvent. Different solvents have been recommended by different workers. Satisfactory results have been obtained using acetone, ethyl alcohol, or methyl alcohol. The alcohols seem to leave smaller amounts of film, but they can dissolve oil from the fingers and deposit that oil on the mirror surface. In addition, the use of pure ethyl alcohol involves governmental restrictions.

To clean the surface, one should hold a piece of lens tissue above the surface and place a few drops of solvent on the tissue, using an eye dropper. Then the lens tissue should be lowered onto the surface and pulled across the surface in one smooth continuous motion. If the surface is not clean, the procedure should be repeated, using a clean piece of lens tissue. The lens tissue is dropped onto the surface and pulled across it so that the dry portion of the tissue removes any solvent residue.

For semihard coatings, one should not use soap and water. If there are large amounts of dirt on the surface, the surface may be flooded with acetone, which may

then be blown off with a jet of dry nitrogen. Then, in the final cleaning step, a piece of lens tissue should be folded several times, and the thick folded edge squirted with a few drops of solvent. The folded edge is then wiped across the surface in one smooth, continuous stroke. If the surface is not clean, this step may be repeated, using a new piece of lens tissue.

Metallic surfaces are sometimes used as the high-reflectivity mirrors in CO_2 lasers. Copper and gold surfaces are soft and may be easily damaged by contact. For such mirrors, one should use noncontact flushing with acetone. Then, the acetone may be blown off with dry nitrogen.

If older soft coatings are to be cleaned, they may be damaged by any solvent. Such mirrors have not been commonly used since the mid 1960s. However, if an older laser with a mirror with a soft coating needs cleaning, perhaps the best method is to blow any contamination off the surface with a jet of dry nitrogen.

As an alternative to dielectric mirrors, some other types of mirror structures are available that offer increased laser damage threshold and that may be suitable for some applications. We shall discuss two alternative types of reflectors: pellicles and etalons.

Pellicles are thin tough films of organic material, like nitrocellulose, which may replace dielectric coatings in some beam splitter applications. Pellicles are relatively low in cost. Because the thickness of the pellicle may be of the order of micrometers, there is virtually no displacement of the reflections from the front and back surfaces of the pellicle. Pellicles are best suited for use in the wavelength region from 0.4 to 2.5 μm.

An etalon is formed by a stack of transmitting plates of a material with a fairly high index of refraction, such as sapphire. The plates are aligned carefully so that all surfaces are parallel. The reflection occurs through a cooperative interference effect. The reflectivity R that may be obtained is

$$R = \frac{(n^{2N} - 1)^2}{(n^{2N} + 1)^2} \tag{6.2}$$

where n is the index of refraction and N is the number of plates. Table 6-3 gives the value of reflectivity as a function of the number of plates for sapphire plates, with an index of refraction of 1.78. Even a single plate gives a reflection of 27 percent, which may be suitable for the output mirror in a high-power laser. As the number of plates increases, alignment and fabrication problems become difficult, but very high reflectivity may be obtained.

3. GAS LASERS

The active material in a gas laser is not subject to optical damage in the same way that the laser rod in a high-power solid state laser is. Moreover, in most CO_2 lasers the gas is flowed through the laser and exhausted. The problem of gas decomposition in sealed-off CO_2 lasers has been discussed earlier.

The mirrors in high-power gas lasers are subject to the same types of optical damage as the mirrors in solid state lasers. In a laser operating in a middle range of

Table 6-3 **Reflectivity of Sapphire Etalon**

Number of plates	1	2	3	4	5
Reflectivity	0.27	0.67	0.89	0.96	0.99

power and capable of material processing, mirror replacement may be required every few thousand hours.

A limiting factor in many gas lasers is depletion of the gas supply, which causes the laser output to drop after some time. In helium–neon lasers, the time may be many thousands of hours, but in argon lasers it may be shorter. Many commercial argon lasers have refill reservoirs that may be activated as needed to restore the gas pressure and the laser output to their proper values. The refill should be done when the laser is warmed up; if it is done when the laser is cool, after warm-up the pressure may be too high. When the supply of gas has been exhausted, the remedy is to return the tube to the manufacturer for refill.

4. FLASHTUBES AND ARC LAMPS

Careful handling is important to ensure the maximum life for flashtubes and arc lamps. The lamps should not be stressed mechanically; stress may damage the seals. They should be handled gently to avoid chipping or scratching the envelope. The lamps must be kept clean; this includes avoiding contamination of the surface by fingerprints. Lamps may be cleaned by careful wiping with lint-free tissue soaked in methanol.

To derive long life from flashtubes, they should be operated at a small fraction (10 to 20 percent) of their single-shot explosion energy. This has been discussed in more detail in an earlier section.

The useful life of both flashtubes and arc lamps increases rapidly as their power loading is decreased. For economical operation and long lamp life, one should design or choose a laser that may be operated at low values of flashtube or arc lamp loading and still produce the desired output power.

5. SEMICONDUCTOR LASERS

The most important method available to the user to extend the life of semiconductor lasers is to make sure that the laser is not driven at too high a current. The user should limit the current to the maximum value specified by the manufacturer for the specific laser. It is desirable that the power supply be designed so that the maximum current cannot be exceeded.

Because unexpected power supply transients can "electrocute" a semiconductor laser without warning, one should use power supplies specifically designed for use with laser diodes. These power supplies have slow start-up features that reduce initial current transients. During start-up, the laser should be connected to the power supply with the supply on and the voltage set to zero. Then, the current may then be

increased slowly to the specified value. The current and light output should be monitored during this procedure. During turn-off this procedure may be reversed.

To reduce the possibility of static electrical discharges, the user should wear grounded wrist straps and grounded foot straps when touching semiconductor lasers. Work surfaces should also be grounded. When not in use, a diode laser should be stored in its antistatic shipping bag.

Because the lifetime decreases exponentially with increasing temperature, the laser must be in good thermal contact with a heat sink capable of dissipating the thermal load generated by the laser. Manufacturers recommend that the heat sink have a thermal impedance less than 2°C/W.

6. LASER POWER SUPPLIES

The high-voltage power supplies used with lasers are subject to degradation, particularly if there is a pulsed electrical discharge. Electrode material is sputtered, and the electrodes become pitted. The variety of possible power supply designs precludes all but the most general of comments. Electrode surfaces must periodically be inspected and reconditioned in order to optimize operation of lasers that require high current flow. An example is CO_2 TEA lasers, for which occasional cleaning and polishing of pitted electrode surfaces may be needed.

7. OPERATING COSTS

All lasers have some operating costs. These include expendables, such as electricity, coolants, and gases, as well as repair and maintenance of components that degrade or suffer damage. The cost may be low, perhaps a few pennies per hour, for low-power He–Ne lasers. For lasers with modest material processing capabilities, such as CO_2 or Nd:YAG, operating costs may be several dollars per hour. Very high power lasers, especially those capable of causing optical damage to their components, can be very expensive to operate.

Whatever the application, the user should estimate the costs for the laser system. The cost should include the care and maintenance required, the possible periodic replacement of degradable items, and any expendable materials. This procedure is essential in the economic analysis of any laser application in industry.

References

[1] D. W. Fradin, E. Yablonovitch, and M. Bass, *Appl. Opt.* **12**, 700 (1973).
[2] A. Vaidyanathan, T. W. Walker, and A. H. Guenther, *IEEE J. Quantum Electron.* **QE-16**, 89 (1980).
[3] N. Bloembergen, *Appl. Opt.* **12**, 661 (1973).
[4] J. Becker and V. Scheuer, *Appl. Opt.* **29**, 4303 (1990).
[5] T. T. Saito, G. B. Charlton, and J. S. Loomis, in *Laser-Induced Damage in Optical Materials*, 1974 (A. J. Glass and A. H. Guenther, eds.), NBS Special Publication 414, U. S. Dept. of Commerce, Washington, D.C., 1974, p. 103.

Selected Additional References

A. Damage and Deterioration of Lasers

J. W. Arenberg *et al.*, Correlating Laser Damage Tests, *Appl. Opt.* **28**, 123 (1989).

H. E. Bennett, *et al.*, eds., *Laser-Induced Damage in Optical Materials: 1991*, *SPIE Proceedings*, Vol. 1624, SPIE, Bellingham, WA, 1992. This is the proceedings of a conference (the so-called Laser Damage Conference) held annually in Boulder, CO, and is one of an extensive series of proceedings of this conference.

W. J. Fritz, Analysis of Rapid Degradation in High-Power (AlGa)As Laser Diodes, *IEEE J. Quantum Electron.* **QE-26**, 68 (1990).

H. Kressel and J. K. Butler, *Semiconductor Lasers and Heterojunction LEDs*, Academic Press, New York, 1977, Chapter 16.

J. F. Ready, *Effects of High-Power Laser Radiation*, Academic Press, New York, 1971, Chapter 6.

S. E. Watkins *et al.*, Electrical Performance of Laser Damaged Silicon Photodiodes, *Appl. Opt.* **29**, 827 (1990)

R. M. Wood, *Laser Damage in Optical Materials*, Adam Hilger, Bristol and Boston, 1986.

B. Care and Maintenance

M. Fukuda, *Reliability and Degradation of Semiconductor Lasers and LEDs*, Artech House, Boston, 1991.

T. R. Gagnier, Calculations of laser reliability can aid in predicting a system's operating life, *Laser Focus*, p. 86 (October 1976).

J. Littlechild and D. Mossler, Knowledge of Arc-Lamp Aging and Lifetime Effects Can Help to Avoid Unpleasant Surprises, *Laser Focus/Electro-Optics*, p. 67 (November 1988).

Chapter 7 | Laser Safety

There are definite hazards that accompany lasers. The most obvious hazard is the possibility of damage to the eye. Skin burns can also be produced by high-power lasers. The safety issues related to the use of lasers have been the subject of intense study for many years. Laser safety standards have also been developed over a long period of time.

With a well-designed safety program, the potential dangers can be reduced to a minimum for scientific, engineering, and industrial applications. Lasers are not unique with respect to hazards; many other types of industrial equipment have associated dangers. The use of lasers in industry may yield desirable results at a level of risk consistent with the use of other industrial tools.

In this chapter, we first describe some of the physiological effects of laser radiation, with emphasis on the mechanisms through which laser radiation may be harmful to the human body. We then discuss safety measures, emphasizing the approach of the American National Standards Institute (ANSI). We finally describe some of the regulatory standards for laser safety.

We shall not describe medical applications of lasers. Although such applications, which include treatment of diabetic retinopathy, attachment of detached retinas, removal of tattoos, and laser angioplasty, are surely important, they do not fall within the scope of industrial applications.

In what follows, we shall emphasize the very real hazards of laser radiation. We should not forget that laser light is light, and human beings have always been exposed to light. We are not dealing with a new agent the effects of which are unknown and mysterious. Laser light, because of its unusual properties, can be focused to a small spot on the retina and thus may be a greater eye hazard than a nonlaser source of the same power. Basically, though, the hazards from laser light are not qualitatively different from those of other high-intensity light sources.

A. Physiological Effects

The structures of the body most at risk for laser light are the retina (the photosensitive surface at the back of the eyeball), the cornea (the front transparent layer of the eye), and the skin. The retina can be damaged by light from visible ($0.4 < \lambda < 0.7$ μm) and near infrared ($0.7 < \lambda < 1.4$ μm) lasers. The light from ultraviolet ($\lambda < 0.4$ μm) and far infrared ($\lambda > 1.4$ μm) lasers does not reach the retina, but it can harm the cornea. The skin can be affected by lasers of any wavelength.

1. EFFECTS ON THE EYE

Accidental exposure of the eye during alignment of a laser beam is the most common cause of laser accidents [1].

Only light in the wavelength range 0.4 to 1.4 μm can penetrate through the anterior structures of the eye and reach the retina. Figure 7-1 shows one curve representing the transmission of light through the front structures of the eye and reaching the retina, and a second curve representing the product of the transmission through the ocular media and the absorption in the various layers of the retina. The second curve

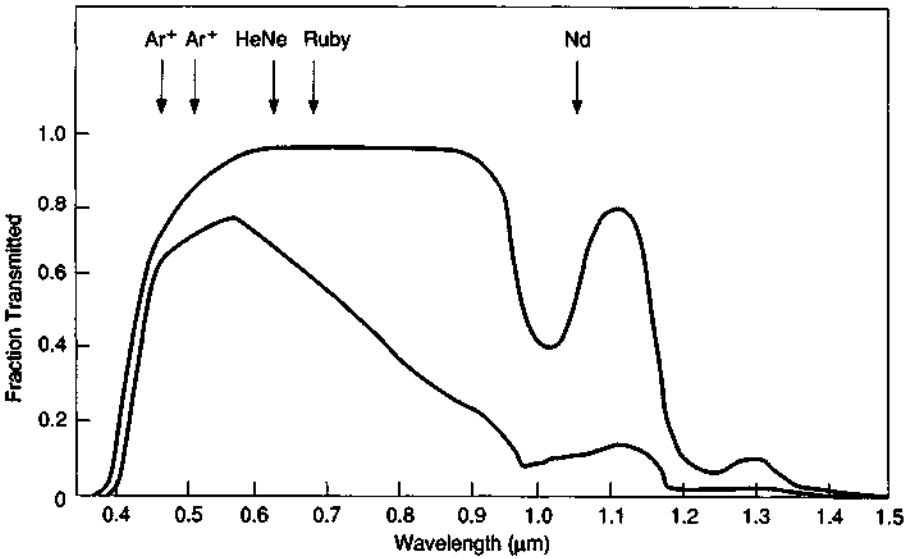


Figure 7-1 Spectral characteristics of the human eye. The top curve shows the percentage transmission through the anterior structures of the eye, i.e., the fraction of the incident light that reaches the retina. The lower curve shows the product of this transmission and the absorption in the various layers of the retina; it gives the fraction of the incident light that is absorbed in the retina. The lower curve defines the relative hazard for the retina. The wavelengths of some common lasers are noted. (From W. J. Geeraets and E. R. Berry, *Am. J. Ophthal.* **66**, 15 (1968).)

essentially defines the hazard to the retina as a function of wavelength. The wavelengths of some common lasers are indicated for reference. Light of wavelength shorter than 0.4 μm or longer than 1.4 μm will not affect the retina. Of course, if the light is intense enough, it can damage the structures in which it is absorbed, usually the cornea. For example, a CO_2 laser is a hazard primarily to the cornea.

The retina is the organ most at risk, because of the focusing action of the lens of the eye. A collimated beam of light entering the eye may be focused to a small area on the retina, so that the irradiance at the focus may be much higher than in the incident beam. Because the diameter of the focal spot at the retina may be as small as 20 μm, a very large increase of the irradiance is possible. The appropriate factor for enhancement of irradiance is $(d_p/d_r)^2$, where d_p is the diameter of the pupil of the eye and d_r is the diameter at the retina. A representative case might be $d_p \approx 0.5$ cm and $d_r \approx 0.002$ cm. This would yield an irradiance at the retina about 6×10^4 times larger than the irradiance entering the eye. Thus, the retina may be damaged at levels where other structures of the body would not be endangered.

Figure 7-2 shows some data on the laser irradiance required to form minimal lesions (burns) on the retina, as a function of duration of exposure. These data were obtained for rabbit eyes. The results express the experimentally determined threshold required to form a minimal barely observable burn on the retina in terms of irradiance *at the retina*. This threshold is the value at which the observers could barely detect small burns when they observed the retina through an ophthalmoscope. Of

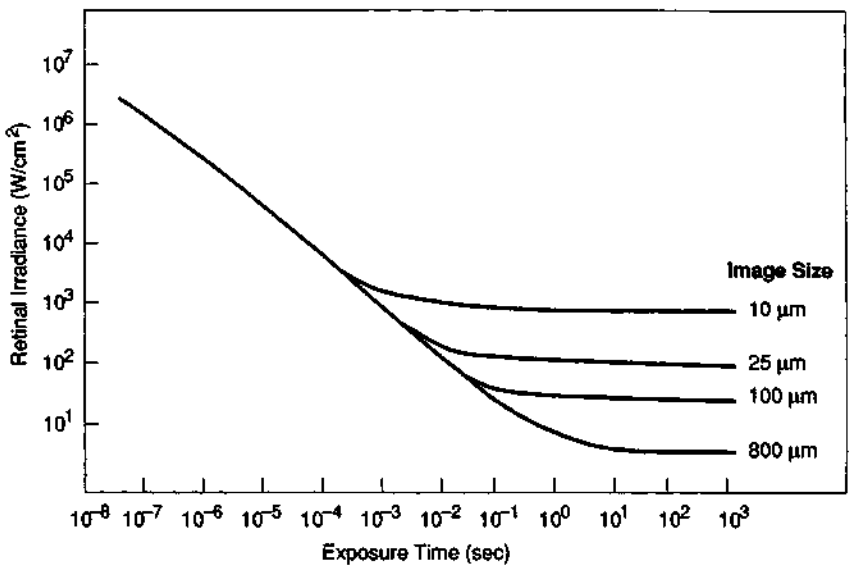


Figure 7-2 Threshold irradiance to cause minimum ophthalmologically observable lesions on the retina of rabbit eyes, for retinal image sizes as indicated and as a function of exposure duration. (From A. M. Clarke *et al.*, *Arch. Environ. Health* **18**, 424 (1969).)

course, at higher values of irradiance, more serious damage occurs, including production of large burns, craters in the retina, ejection of material into the vitreous medium of the eye, and hemorrhage within the eyeball. The figure shows that the threshold irradiance is a strong function of both exposure duration and retinal image size.

The dependence of the threshold on the diameter of the focal spot may be understood in terms of thermal conduction. For very short exposure times, there is very little thermal conduction out of the spot, and the threshold is independent of spot size. For longer exposure times, there is conduction of heat energy out of the spot during the exposure, and a higher irradiance is required to form a lesion in a small spot than in a large spot, because the thermal gradients will be higher in a small spot.

In addition, at short exposure duration, where thermal conduction is unimportant, the threshold irradiance scales approximately as the inverse of the duration, suggesting that a threshold fluence, and a threshold temperature rise, are required for damage. The results are consistent with a purely thermal origin for threshold retinal burns.

The data in Figure 7-2 do not completely define the hazard of laser radiation for the retina. These data were obtained mainly with light sources near the red end of the visible spectrum. Sources near the blue end may have more damaging effects, as will be described later. Further, the extrapolation of results from rabbit eyes to human eyes is not certain. The effects of cumulative exposures to repeated subthreshold pulses are not completely understood. Still, there are enough data to allow establishment of safety guidelines. Results such as those shown in Figure 7-2 have been employed in the establishment of laser safety standards.

The threshold values in Figure 7-2 may easily be exceeded by commonly available lasers. Many lasers are capable of delivering potentially damaging exposures to the retina, even at large distances from the laser. High-power pulsed lasers can deliver values of irradiance to the retina much in excess of the thresholds in the figure, even at distances of many miles from the laser.

The results expressed in Figure 7-2 appear to be consistent with a thermal model, in which the damage is produced by heating effects. Later work with lasers operating at wavelengths throughout the visible spectrum indicated that additional effects, possibly photochemical effects, are needed to account for results near the blue end of the visible spectrum. Figure 7-3 shows the wavelength dependence. The figure plots the reciprocal of the threshold retinal irradiance required to produce a minimum retinal burn in an anesthetized rhesus monkey, for several exposure times. The variation cannot be explained by a purely thermal model, and it appears that photochemical action causes the threshold to decrease at shorter wavelengths. This means that for equal power, a blue laser will be more dangerous to the retina than a red laser.

The data in Figures 7-2 and 7-3 express results in terms of irradiance at the retina. This is difficult to characterize in practical situations because of the variable focusing of the lens of the eye. For a given power entering the eye, the irradiance at the retina may be difficult to predict. It is easier to characterize the total power entering

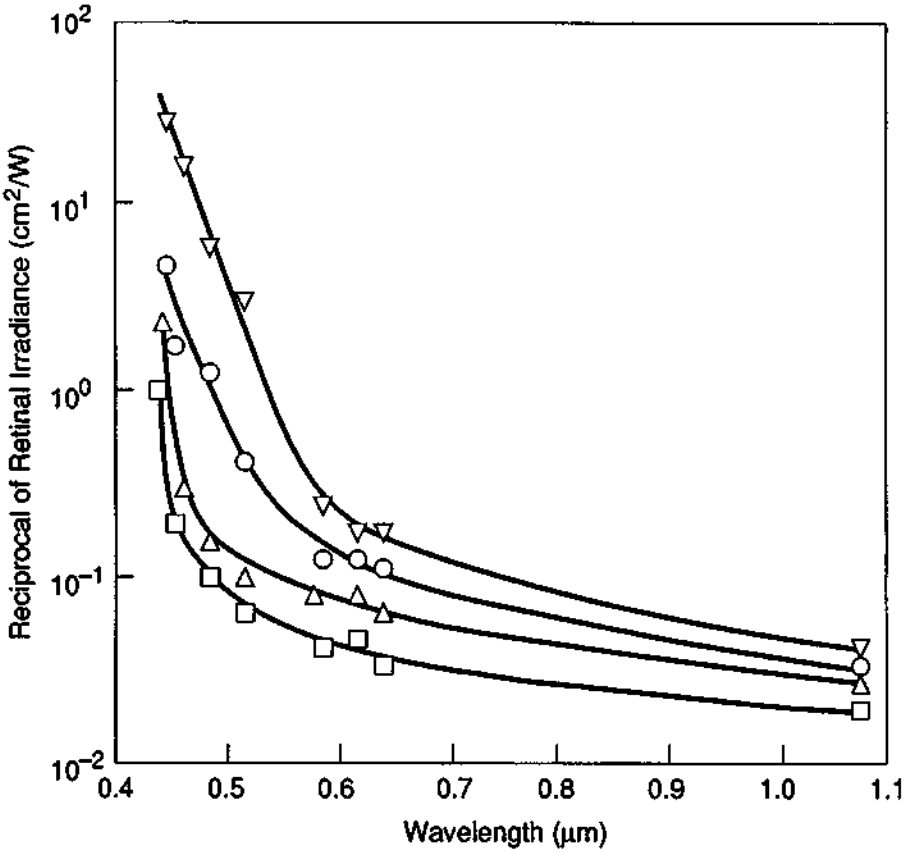


Figure 7-3 Retinal sensitivity to threshold damage in the rhesus monkey, as a function of wavelength. The ordinate gives the reciprocal of the irradiance needed to cause a threshold lesion. Pulse durations: □, 1 sec; △, 16 sec; ○, 100 sec; ▽, 1000 sec. (From W. T. Ham, H. A. Mueller, and D. Sliney, *Nature* **260**, 153 (1976).)

the eye. Figure 7-4 shows the threshold for minimal lesions on the retinas of monkeys caused by exposure to continuous He–Ne laser light, as a function of exposure time. These lasers are of special interest because of their widespread use. Results are shown for irradiation of the macula, the small sensitive region in which vision is most acute, and for areas outside the macula. The minimum power found to cause a lesion was 7 mW, for exposures several seconds long. The monkeys' eyes were rigidly restrained. Over a period of several seconds, it is likely that an accidental human exposure would cause less effect because of the aversion response to bright light.

Light at wavelengths outside the 0.4–1.4 μm range is absorbed in the structures of the eye in front of the retina and is a hazard to those structures. Ultraviolet light (wavelengths less than 0.4 μm) is absorbed in the cornea. Threshold burns to the

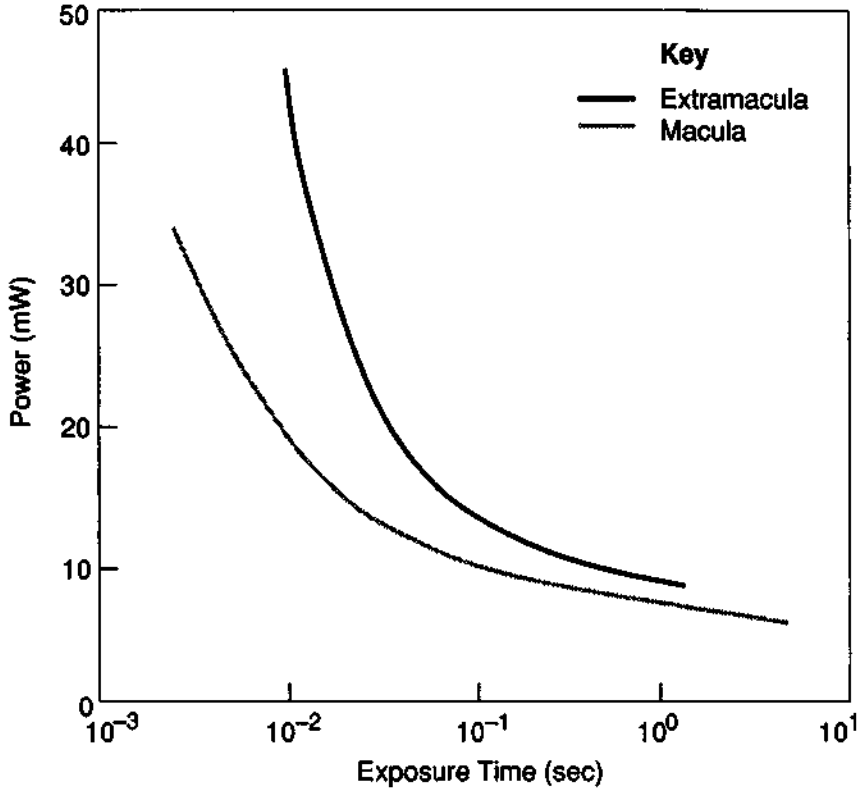


Figure 7-4 Threshold values of He–Ne laser power needed to cause minimal retinal lesions in rhesus monkeys, as a function of exposure duration. The ordinate gives the total power entering the eye. (From P. W. Lappin, *Arch. Environ. Health* **20**, 177 (1970).)

cornea cause reddening and pain similar to "welder's flash." At least some of the damage for ultraviolet burns appears to be photochemical in origin.

For longer infrared wavelengths (>1.8 μm) the light is also absorbed at the cornea, and it can cause corneal burns if the intensity is high enough. The CO_2 laser operating at 10.6 μm is a case of special interest. Some data on the threshold CO_2 laser irradiance required to cause corneal burns in rhesus monkeys are shown in Figure 7-5. We note in comparing Figures 7-2 and 7-5 that the irradiance in Figure 7-2 is defined at the retina and may be larger than that at the cornea by a factor as large as 10^5. Thus, for equal total power in the beam, a visible laser is likely to be more dangerous than a far infrared laser because of the focusing action of the eye.

Light in the wavelength range from 1.53 to 1.8 μm is apparently absorbed in the bulk of the eyeball, rather than at the surface of the cornea or the retina. Thus, the energy is absorbed within a larger volume of material and is less likely to produce damage. This wavelength region is characterized by higher threshold values of irra-

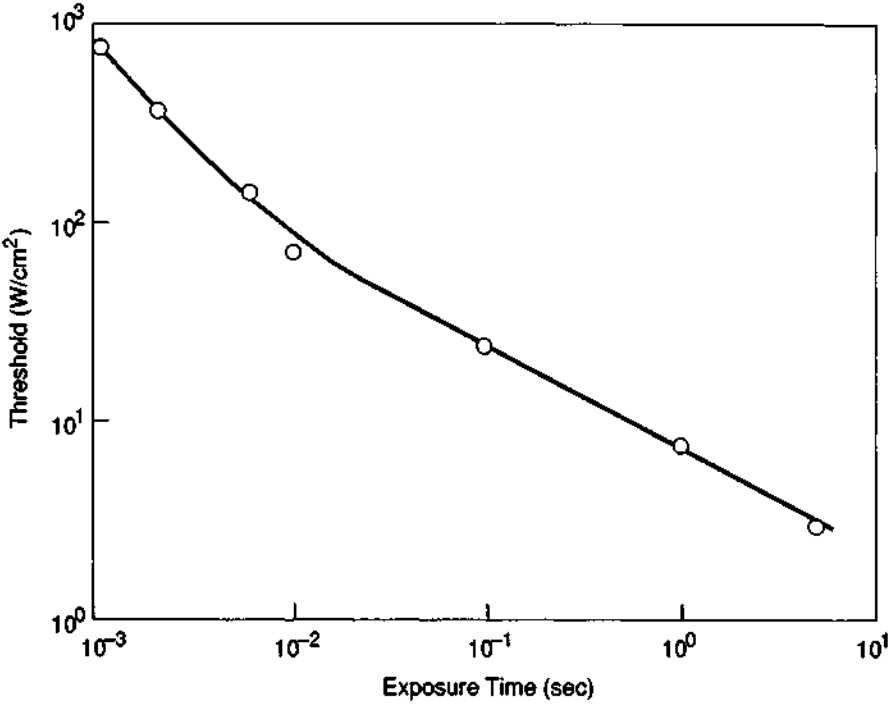


Figure 7-5 Threshold CO_2 laser irradiance needed to cause minimal damage to the cornea of rhesus monkeys as a function of exposure duration. (Data due to B. E. Stuck, Frankford Arsenal.)

diance for burns, and it may represent a region where eye hazards are less acute. There has been considerable effort devoted to the development of lasers operating in this "eye-safe" region, for applications such as range finding.

In this section we have discussed the hazards of laser radiation to the eye in terms of a number of parameters of the exposure, including wavelength, irradiance, and duration. The discussion has quantified the threshold irradiance, for a minimum observable burn in a single exposure, based on data from animals. Some observations have indicated that microscopic histological changes may occur at lower levels than the burn threshold, and so safety standards usually incorporate some reduction below the burn threshold. Moreover, the effect of chronic or multiple exposures is not fully understood. The thresholds are mostly derived from animal experiments, but the small amount of data from human exposures (accidental exposures and volunteers) has indicated that the threshold values are relevant for the establishment of safety standards.

In terms of the functioning of the eye as part of a living human organism, the effect of a small burn on the retina will depend greatly on its location. A small burn on the macula, the small sensitive region in which vision is most acute, could severely

reduce visual acuity. The same small burn on the periphery of the retina would have less noticeable effects. Small retinal burns show at least partial healing over a period of time, with some recovery of the loss of visual acuity.

2. *EFFECTS ON THE SKIN*

High-power lasers can also inflict skin burns. The problem is less acute than for the eye because there is not the increase in irradiance associated with the focusing action of the eye, and because a small burn on another part of the body is likely to be of less concern than a similar burn on the retina or cornea. Still, laser skin burns are a subject of valid concern. High-power lasers in use in industry are potentially capable of inflicting frightful burns.

The effect of laser irradiation of the skin depends on both the wavelength and the pigmentation of the skin. Figure 7-6 shows the reflectivity of the skin as a function of wavelength. In the visible, the skin is quite reflective, with the exact value depending on the pigmentation. Thus, much of the visible light incident on the skin will be reflected. In the far infrared, the skin becomes highly absorbing. The reflectivity is low, independent of pigmentation.

For the light that is not reflected, there is a substantial variation in its transmission through human tissue. A transmission maximum exists near 1.15 μm; nearly 20 percent of the light at this wavelength will be transmitted through the human

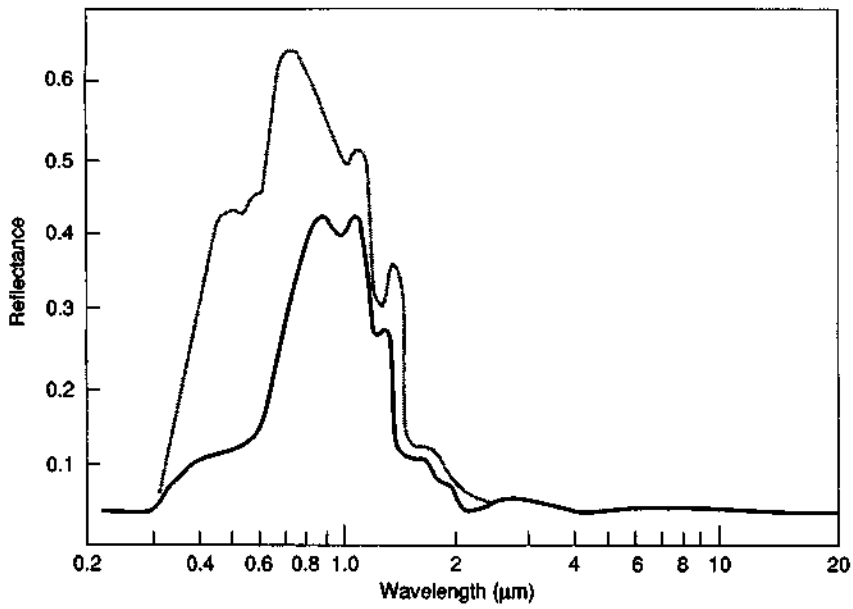


Figure 7-6 Reflectivity of human skin as a function of wavelength. The top curve is for lightly pigmented skin. The lower curve is for heavily pigmented skin. (From D. H. Sliney and B. C. Freasier, *Appl. Opt.* **12**, 1 (1973).)

cheek [2]. Transmission drops as one moves away from this wavelength; at wavelengths greater than 1.4 μm and less than 0.5 μm, transmission through the cheek is virtually zero. At longer infrared wavelengths, in particular at 10 μm, absorption is very high, and it occurs in a thin surface layer.

Thus, CO_2 lasers are of special concern for skin burns. The reason is threefold:

1. CO_2 lasers are the most common type of high-power lasers that can emit at levels high enough to cause severe skin burns.
2. Reflectivity of the skin is low at 10 μm, so that almost all the energy is absorbed, whereas in the visible and near infrared a significant fraction can be reflected.
3. The absorption coefficient is very high, so that the energy is deposited in a thin layer. In contrast, for visible and near infrared lasers, the absorption occurs over a more extended depth, so that heating effects will be less severe.

This discussion does not imply that other lasers cannot be hazards for skin burns, but CO_2 lasers are probably the most significant industrial hazard. For large CO_2 lasers used in industry, even a transient passage through the beam could produce severe burns.

The threshold level at which a CO_2 laser can produce minimal skin burns is shown in Figure 7-7. These data are relevant to the shaved skin of anesthetized white pigs. The threshold criterion as given in this figure was a mild reddening 24 hours after exposure.

The results in Figure 7-7 for threshold levels for skin burns are similar to the threshold levels in Figure 7-5 for threshold corneal damage. The cornea is not particularly more sensitive to damage by a CO_2 laser than are other portions of the skin. It is apparent that a small burn on the cornea may be more troublesome than a similar small burn elsewhere on the skin.

Near the threshold for skin burns, thermal mechanisms appear to dominate and the damage appears to arise from thermal effects. At high levels, well above threshold, other effects occur. For example, the rapid absorption of energy can produce shock waves that travel through the body and can cause damage to tissue removed from the original impact point. Adoption of a safety program that eliminates threshold thermal burns will automatically eliminate damage due to effects associated with higher levels of power.

3. OTHER HAZARDS

Most emphasis in laser safety has been put on the beam emitted by the laser. But there are other possible hazards associated with lasers, for example, electrical shock, poisonous or corrosive substances used in the laser and in associated equipment, possible carcinogens used in dye lasers, and possible noxious vapors produced in laser evaporation of target materials. Probably the most hazardous aspect of a laser is the electrical power supply. A number of people have been electrocuted by contact with laser power supplies. Electrical hazards are not unique to lasers. Precautions similar to those used with other high-voltage electrical equipment are appropriate.

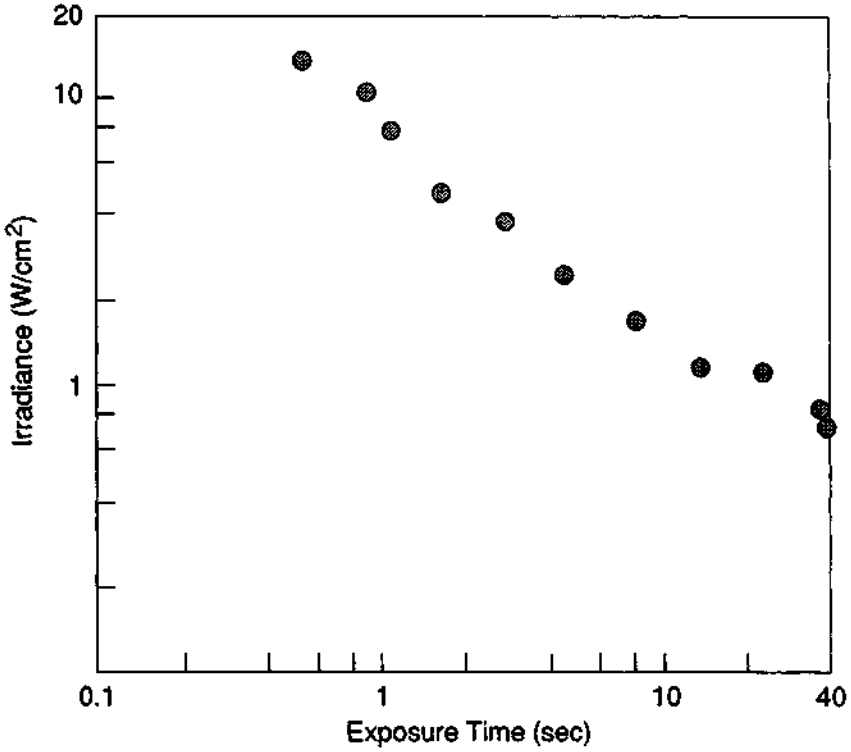


Figure 7-7 Threshold level for minimum skin burns on the skin of white pigs as a function of exposure time, for CO_2 laser radiation. (Data from A. S. Brownell and B. E. Stuck, Frankford Arsenal, AD Report 785.609 (1972).)

Arc lamps used to pump continuous solid state lasers represent a possible explosion hazard because they are pressurized to above atmospheric pressure. Arc lamps can explode if the seals are mechanically stressed during handling or during operation, with resultant flying glass fragments. One should use safety goggles when installing, removing, or cleaning arc lamps and use great caution to avoid mechanically stressing them or scratching them. They should be operated only within their protective housings.

A complete laser safety program must consider not only the hazards associated with the beam, but also the ancillary hazards, including those we have just mentioned.

B. Laser Safety Practices and Standards

In this section, we will emphasize safety practices relevant to the laser beam itself, this being the hazard unique to the use of lasers. The fundamental laser safety practice is not to allow laser radiation at a level higher than a maximum permissible

exposure to strike the body. Maximum permissible exposures have been defined in laser safety standards, as we shall discuss later, for both eyes and skin, and for various wavelengths and exposure durations.

In discussing protection of the eye from the laser beam, we distinguish three cases. First, it is possible to place the eye in the beam path, so that the beam enters the eye directly. Second, it is possible for a specularly reflected beam, that is, a beam reflected from a shiny object like a mirror, to enter the eye. A specularly reflected beam is still collimated and may be as hazardous as the direct beam. Finally, diffusely reflected light, like light reflected from a wall with a flat paint coating, may enter the eye. Such light is no longer well collimated, and it will usually be less hazardous than the direct beam or a specularly reflected beam; but at high laser power, even diffusely reflected light may be hazardous.

Safety measures may include physical barriers and engineering controls. Physical barriers are materials opaque at the laser wavelength that are positioned so that the laser beam cannot strike the body. Engineering controls are defined procedures that ensure that the laser is not activated until the beam path is clear. Both types of control measures have application. Physical barriers may be most appropriate in a production environment. An entire workstation may be enclosed so that the beam cannot escape. Engineering controls may be more appropriate for scientific measurements, where the workers may be more skilled in good optical practices, and where a rigid enclosure might hamper the measurement.

The most common measure for eye protection is to use protective eyewear of sufficient optical density to reduce the exposure to the eye to below the maximum permissible level. Laser safety eyewear is specified by values of the optical density at the laser wavelengths for which it is to be used. Optical density (O.D.) is defined by

$$\text{O.D.} = -\log_{10}\frac{I}{I_0} \tag{7.1}$$

where I_0 is the incident intensity and I is the transmitted intensity. Thus, an optical density of 5 means an attenuation by a factor of 10^5.

Safety eyewear should be employed whenever the possibility exists for exposures above the maximum permissible level. For industrial applications, perhaps the best rule is never to allow direct viewing of the beam, at least for pulsed lasers and for continuous lasers operating above some level.

For pulsed lasers and for high-power continuous lasers, the beam path should be designed so that the operator's eyes cannot get into the beam, and so that there are no specular reflections of the beam outside the protected area. Protective eyewear would be required in cases where diffuse illumination could be a hazard.

There is no one material that provides effective protection for all laser wavelengths and still allows effective vision. There is laser protective eyewear effective for a laser of any particular wavelength, and a number of commercial suppliers offer such eyewear. Ordinary safety eyeglasses with side shields will attenuate CO_2 laser beams considerably, but there may be danger of the glass cracking in high-power beams.

The luminous transmission of the eyewear is an important factor. If the luminous transmission is low, the user may not find the eyewear acceptable.

Laser safety eyewear is available from many commercial suppliers. It is available both as spectacles and as goggles. Goggles offer more complete protection, with no possibility of light entering the eye around the edge. Goggles do cut off side vision and may sometimes be unacceptably hot or heavy. Spectacles with side shields are lighter and probably more acceptable to the wearer than are goggles. People who wear glasses may wear goggles over their glasses or obtain safety eyewear ground to their prescription.

The user should not look directly into the laser beam even when wearing laser safety eyewear. Exposure of laser safety eyewear to high levels of laser light can cause damage in the form of melting, bleaching, cracking, or shattering. If this occurs, the beam can then enter the eye. Some eyewear showed structural changes after several seconds of exposure to laser beams with irradiance around 6–12 W/cm^2 [3]. In some cases a portion of the incident beam was transmitted through holes produced in the eyewear.

The preceding discussion has applied to direct exposure of the eye to laser radiation. The reflected laser beam may also be hazardous, especially because reflections sometimes occur in unexpected directions. Specular reflection from a shiny surface is especially dangerous. Reflections may occur from many types of objects in the beam path: glass surfaces, mirrors, rings, wristwatches, and so forth. In metalworking, one often deliberately places a shiny metal surface in the beam. A specular reflection preserves the collimated nature of the beam and may be as dangerous as the direct beam. The same precautions apply to specularly reflected beams as to the direct beam.

Diffuse reflections may be less dangerous than specular reflections, because the collimation is not preserved. For diffuse reflections, one must consider the irradiance at the retina. For distances of the order of room dimensions, the irradiance at the retina is independent of distance from the reflection. This fact arises because the irradiance at the cornea decreases with the distance from the diffuse reflection, but the focal area on the retina decreases with distance in the same ratio. This makes a maximum permissible exposure at the cornea harder to specify for diffuse reflection than for direct illumination or for specular reflection.

Although a diffuse reflection is less a hazard than a specular reflection, high-power lasers can easily reach levels where the reflection from a diffusely reflecting surface can be hazardous. The dangers from diffusely reflected laser light cannot be ignored when considering appropriate safety measures.

Skin burns are perhaps less of a problem that eye damage. Many common lasers (like the frequently used helium–neon laser) are not capable of producing skin burns. Larger lasers used for metalworking could produce terrible burns, but such installations are relatively easily enclosed. The most appropriate measure for a laser that presents a real hazard for skin burns could be to use physical barriers to separate completely the beam from people.

Operation of high-power lasers in closed areas with restricted access is desirable. Warning lights energized when the laser is activated, suitable warning signs, audible

signals when appropriate, and closed doors with interlocks to the laser power supply are all desirable features.

A laser installation should have a designated safety officer responsible for evaluation and control of potential hazards. Periodic eye examinations for laser personnel are to be recommended. Many institutions provide eye examinations for their laser personnel on a routine basis.

Finally, safety training for personnel who will operate lasers is very desirable. Comprehensive laser training programs for industrial personnel have been developed.

1. THE ANSI STANDARD

Various standards and guidelines have been developed to define laser safety practices. One standard that has particular significance is that developed by the Z-136 Committee of the American National Standards Institute (ANSI) [4]. This voluntary set of standards is perhaps the most influential and widely adopted laser safety standard. The standard is concerned with several areas, including

- Definition of maximum permissible exposure (MPE) levels to laser light
- Classification of lasers into classes according to level of hazard
- Definition of safety practices for each class

The maximum permissible exposures are not meant to be fine lines between danger and safety; rather, they are evaluations of the current status of knowledge about the thresholds at which laser light begins to harm the human body. They represent levels below which no harmful effects are known to occur.

Maximum permissible exposures for the eye are specified at the cornea, rather than at the retina. This is done because it is much easier to measure irradiance at the cornea.

As an example of the MPE levels defined in the ANSI standard, Figure 7-8 shows the ocular MPE for intrabeam viewing as a function of exposure duration and wavelength in the visible and near infrared portions of the spectrum for beams that subtend small angles. The MPE is expressed in terms of irradiance at the cornea, and it is relevant for single exposures to a collimated laser beam. The curves in Figure 7-8 cover only a small fraction of the cases defined in the ANSI document. Other definitions are relevant for extended sources and for other wavelength regions.

This figure defines the MPE for some commonly encountered lasers. For wavelengths between 0.7 and 1.051 μm, and between 0.55 and 0.7 μm, correction factors are defined that allow determination of the MPE for any wavelength in those regions.

According to the figure, the MPE is reduced for some values of exposure duration for the shorter-wavelength regions of the visible spectrum, with the amount of reduction becoming greater as the wavelength moves farther toward the violet end of the visible spectrum. The reduction of the MPE at shorter wavelengths is influenced by research results indicating that photochemical mechanisms may begin to influence retinal damage at short visible wavelengths.

The values of MPE in Figure 7-8 apply to single exposures. For multiple exposures, the MPE for wavelengths greater than 700 nm is reduced by a factor of $n^{-0.25}$,

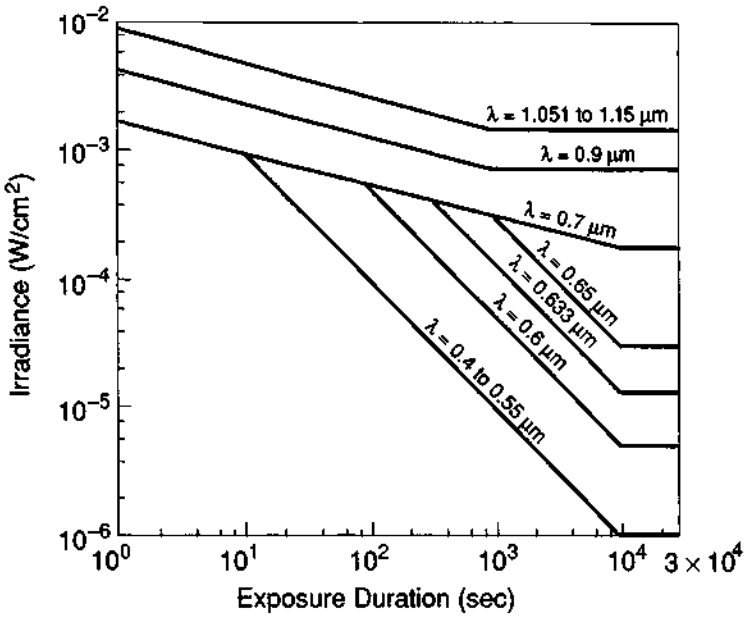


Figure 7-8 Maximum permissible exposure for intrabeam viewing as a function of exposure duration at the specified wavelengths. (From American National Standard for Safe Use of Lasers, ANSI Z136.1 (1993), published by the Laser Institute of America, Orlando, FL.)

where n is the number of exposures. For wavelengths less than 700 nm, the MPE is reduced by the same factor, but it also must not exceed the MPE calculated for nt seconds, when nt is greater than 10 sec, with t the pulse duration.

The ANSI standard also defines MPE values for exposure of the skin. These values depend on wavelength, with especially strong dependence in the ultraviolet region from 0.303 to 0.315 μm, and, in some cases, on exposure duration.

A second important topic in the ANSI standard is the grouping of lasers into four classes, according to level of hazard. The philosophy of the four classes appears to be as follows:

Class 1 The emission of power or energy accessible to human exposure is below the levels at which harmful effects are known to occur.

Class 2 Relatively low power *visible* lasers, for which the aversion response will help prevent eye damage. A subclass, 2A, is for lasers for which exposure for a period less than 1000 seconds should not be hazardous, but for which an exposure exceeding that period could be.

Class 3 Medium-power lasers that can emit at levels such that exposure to the direct beam would be harmful, but for which diffuse reflections should not be harmful. This class is divided into two subclasses, depending on whether the output is between one and five times the upper limit of the next lower class (class 3A) or above that value (class 3B).

Class 4 High-power lasers that can emit at levels such that harmful effects from diffuse reflections could occur.

The classification implies that, above some level, even diffuse reflections may be hazardous. Figure 7-9 shows values for the maximum laser power at a wavelength of 514 nm incident on a diffusely reflecting surface for which the irradiance at the cornea will not exceed the MPE given in Figure 7-8. The figure assumes a 100 percent diffusely reflecting Lambertian surface viewed normal to the surface at a distance of 1 m, with a laser spot size on the surface very small compared with 1 m. The figure indicates that commonly available lasers could yield diffuse reflections that exceed the MPE for intrabeam viewing. Such reflected light presumably could be hazardous.

The exact classification for a given laser depends on the wavelength and the duration of the possible exposure, and it is somewhat complicated. The full ANSI document should be consulted for details of the classification system. For reference, the classifications are given in Table 7-1 for a few common types of lasers.

The table gives the upper limit for classes 1, 2, and 3. Membership in class 1 is for lasers the emission of which is below the limit for that class. Membership in classes 2 and 3 is for those lasers the emission level of which is above the limit for the next lower class and below the limit for the given class. Class 4 lasers are those with emission levels that exceed the limit for class 3. For continuous lasers (emission duration >0.25 sec), the levels are given in units of power. For pulsed lasers, the limits are given in units either of energy or fluence per pulse. For repetitively pulsed

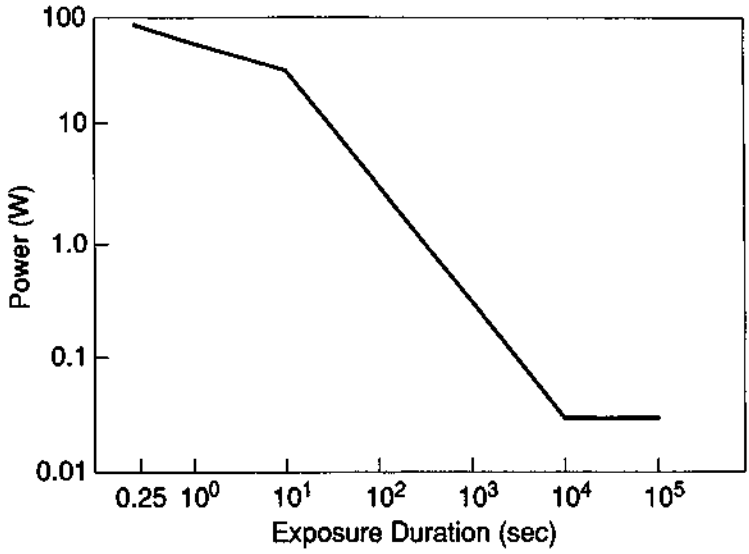


Figure 7-9 Maximum laser power at a wavelength of 514 nm incident on a diffusely reflecting Lambertian surface that will not exceed the MPE for intrabeam viewing when the surface is viewed normal to the surface at a distance of 1 m.

Table 7-1 **Examples of Classification of Common Lasers**

Laser	Wavelength (μm)	Upper limit			
		Class 1	Class 2[a]	Class 3[b]	Class 4
He–Ne (continuous)	0.6328	6.98 μW	1 mW	0.5 W	>0.5 W
Argon (continuous)	0.5145	0.4 μW	1 mW	0.5 W	>0.5 W
CO_2 (continuous)	10.6	0.8 mW	—	0.5 W	>0.5 W
CO_2 (TEA, 100 nsec pulse)	10.6	80 μJ	—	10 J/cm^2	>10 J/cm^2
Nd:YAG (continuous)	1.06	0.2 mW	—	0.5 W	>0.5 W
Nd:YAG (20 nsec pulse)	1.06	2 μJ	—	0.16 J/cm^2	>0.16 J/cm^2

[a]Class 2 is relevant only for visible lasers.

[b]Class 3 is divided into classes 3A and 3B, with class 3A containing those members of class 3 with accessible outputs between one and five times the upper limit for the next lower class, and class 3B containing those members with accessible outputs greater than five times the upper limit for the next lower class.

lasers, the single-pulse energy and the average power must both fall below the upper limit for a given class if the laser is to be in that class.

The ANSI document also describes specific control measures to reduce the possibility of exposure of eyes and skin to hazardous levels of laser radiation. The control measures recommended for each of the four classes of lasers we have described are specified. The control measures include items such as protective eyewear, interlocks, enclosed beam paths, beam stops, use in enclosed areas, equipment labels, and warning signs. The amount of recommended control increases with the degree of hazard for the different laser classes. The recommendations for control measures are quite detailed. In formulating a laser safety program, one should refer to the complete ANSI document.

2. *REGULATORY BODIES*

The ANSI laser safety standard is voluntary. There are a number of bodies that enforce nonvoluntary laser safety standards and practices. Some of these entities are listed in the following:

Employers Many employers enforce laser safety standards for their workers. In many cases, the standard is the ANSI standard.

The Food and Drug Administration (FDA) This is part of the Department of Health, Education, and Welfare. The standards the FDA promulgated became effective in 1976 [5] and have been modified in 1978 and 1985. The FDA regulations affect primarily the manufacturers of laser products (or a buyer who builds the laser into a system for resale), rather than the user. The regulations require the

manufacturer to incorporate specified controls, markings, and warnings into laser products and to classify, test, and certify the laser products.

The Occupational Safety and Health Administration (OSHA) This is part of the Department of Labor. OSHA has not promulgated general standards for the safe use of lasers in industry, with the exception of regulations regarding the use of lasers in construction.

States Many individual states have laws regarding registration and control of lasers. The individual laws vary widely from one state to another. It is imperative that the user determine the relevant state laws.

International bodies The International Electrotechnical Commission (IEC) standard on Radiation Safety of Laser Products, Equipment Classification, Requirements and User's Guide, Publication 825, appears likely to be adopted as the standard for laser safety for the European Community.

References

[1] R. J. Rockwell, *J. Laser Appl.* **1**, 53 (1989).
[2] C. H. Cartwright, *J. Opt. Soc. Am.* **20**, 81 (1930).
[3] K. R. Envall *et al.*, *Preliminary Evaluation of Commercially Available Laser Protective Eyewear*, HEW Publication (FDA) 75-8026, March 1975.
[4] Standard Z136.1 (1993), *Safe Use of Lasers*, Laser Institute of America, Orlando, FL (1993).
[5] United States Code of Federal Regulations, Title 21, Subchapter J.

Selected Additional References

J. E. Dennis and R. G. Bostrom, Reviewing the Federal Standard for Laser Products, *Lasers and Optronics*, p. 95 (March 1988).
J. E. Dennis and D. H. Edmunds, Comparison of the CDRH and IEC Standards for Laser Products, *J. Laser Appl.* **3**, 7 (Fall 1991).
T. J. Glynn and M. Walsh, Laser Safety Primer Outline for Industrial/Medical Users, *J. Laser Appl.* **4**, 33 (Fall 1992).
W. T. Ham, Jr., and H. A. Mueller, ANSI Z136.2 Update: Ocular Effects of Infrared Radiation, *J. Laser Appl.* **3**, 19 (Fall 1991).
A. Kestenbaum, R. J. Coyle, and P. P. Nolan, Safe Laser System Design for Production, *J. Laser Appl.* **7**, 31 (1995).
L. Matthews and G. Garcia, *Laser and Eye Safety in the Laboratory*, SPIE, Bellingham, WA, 1994.
A. F. McKinlay, F. Harlen, and M. J. Whitlock, *Hazards of Optical Radiation*, Adam Hilger, Bristol and Philadelphia, 1988.
R. J. Rockwell *et al.*, ANSI Z136.1 Proposed 1992 Changes, *J. Laser Appl.* **4**, 45 (Winter 1992).
D. Sliney and M. Wolbarsht, *Safety with Lasers and Other Optical Sources: A Comprehensive Handbook*, Plenum Press, New York, 1980.
R. Weiner, Status of Laser Safety Requirements, *Lasers & Optronics*, p. 74 (September 30, 1993).

Chapter 8 | Alignment, Tooling, and Angle Tracking

One of the most common applications of lasers in industry arises in the solution of alignment problems, in which knowledge of the displacement from a reference line or plane is desired. Use of laser alignment techniques has replaced older, more tedious methods of performing such tasks as machine tool alignment. These applications are made possible by the unusual properties of laser radiation. The most important properties are the directionality of the beam and the high radiance.

A collimated Gaussian laser beam has a spatial distribution given by

$$I(r) = I_0 \exp \frac{-2r^2}{r_0^2} \tag{8.1}$$

where r is the radial distance from the beam center, and r_0 is the radius at which the irradiance has fallen to $1/e^2$ of its value I_0 at the center of the beam. The distribution will remain Gaussian at all distances from the laser. For a Gaussian beam, the divergence angle θ is

$$\theta = \frac{2\lambda}{\pi D} \tag{8.2}$$

where λ is the wavelength and D the diameter of the exit aperture of the laser. This defines the half-angle of the cone into which the beam diverges at distances far from the laser. A typical He–Ne laser with an exit diameter around 1 mm will have a beam divergence angle around 10^{-3} rad. With external collimating optics, the beam divergence may be reduced further. The excellent collimation of the beam allows precise positioning of objects at positions remote from the laser.

The radiance is important because the laser beam will be easily visible, even in an ambient background of room light or sunlight. Most practical laser systems for alignment have employed He–Ne lasers, typically at power levels around 1–5 mW. The orange-red beam from such a laser is easy to detect, even in a sunlit background at fairly large distances from the laser. It is usually unnecessary to use lasers with

higher power. The He–Ne laser is reliable and inexpensive. This has made it the laser of choice for alignment applications in which one requires a visible beam with modest power. In recent times, the development of visible semiconductor diode lasers, packaged to yield a collimated beam, is offering an alternative to the He–Ne laser, and this type of laser is beginning to displace the He–Ne laser for some alignment applications.

Because many of the applications require measurement of the position of a laser beam, we begin the chapter with a discussion of position-sensitive detectors. We will then describe the use of lasers for alignment applications, sometimes referred to as "tooling." We will discuss related applications, such as alignment in a plane and the closely associated application of angle tracking. We will also summarize some of the ways in which these techniques have been employed in the construction industry.

A. Position-Sensitive Detectors

There are several detector types that can measure the position of a laser beam that strikes the detector. They include

- Quadrant photodiodes
- Lateral-effect photodiodes
- Photodiode arrays
- Image dissector tubes
- Vidicons

These devices are often used in alignment and tracking applications.

The quadrant photodiode uses four silicon photodiodes, arranged as shown schematically in Figure 8-1. This configuration is also called a centering detector. The photodiodes are electronically connected in quadrature to compare the intensity in each half of the beam, both horizontally and vertically. If the beam accurately strikes the center of the detector, each quadrant receives the same amount of light. One then observes a null signal between the pairs of photodiodes. If the beam is displaced from the center of the detector, there will be a different amount of light incident on each detector, and the balance in the signals between pairs of detectors will no longer hold. The output may be presented on a meter. A null signal on the meter indicates that the beam is accurately centered on the quadrant detector. Deviation from a null quantitatively indicates the displacement of the beam. The sign of the signal indicates the direction of the displacement.

For proper operation of the quadrant detector, the laser beam should be Gaussian. If admixtures of higher-order modes are present, the beam profile may not be symmetric, and the measurement may be complicated.

Another type of position-sensitive detector is the lateral-effect photodiode. This is a PIN silicon photodiode (see Chapter 5) with metallic contacts separated by

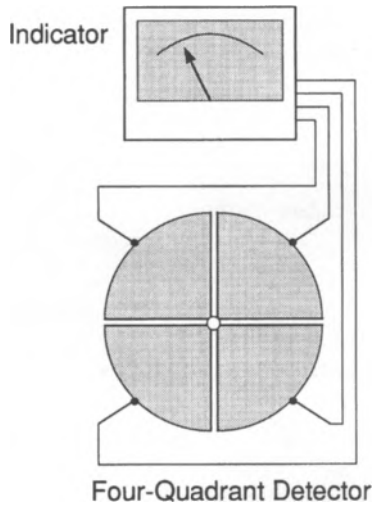


Figure 8-1 Centering detector consisting of four silicon photodiodes.

distance L. As a spot of light is scanned across the interelectrode spacing, the fraction of the current carried by each of the electrodes varies. The current I carried by one of the electrodes when the spot of light is at a distance x from that electrode is

$$I = I_0 \frac{\sinh \alpha(L - x)}{\sinh \alpha L} \tag{8.3}$$

where I_0 is the total photocurrent and α is a material parameter. When $x = 0$, $I = I_0$, and all the current is carried by the specified electrode. When $x = L$, $I = 0$, and all the current is carried by the other electrode. Thus, monitoring the current carried by each electrode can provide a direct measurement of the displacement of the spot of light across the active area of the detector.

Arrays of silicon photodiodes can provide a simple indication of the location of the laser beam through observation of which diodes are illuminated. Two-dimensional arrays with hundreds of diodes in each dimension are available. The resolution of the array is equal to the spacing between detectors.

An image dissector is a modified photomultiplier tube (see Chapter 5). The modifications that make a photomultiplier an image dissector include an electrostatic field that focuses the photoelectrons from the photocathode, an aperture for the electron beam, and a deflection scheme that allows one to steer the beam through the aperture. Knowledge of the deflection required to steer the beam through the aperture then identifies the position of the image on the photocathode.

A vidicon is a type of imaging detector used in television cameras. Vidicons are used as position-sensitive detectors in a manner similar to the image dissector. Position information is obtained from the deflection current that is applied when a signal is detected. Internally, the vidicon is fundamentally different from the image dissec-

tor. The photosensitive surface is photoconductive rather than photoemissive, and the vidicon does not have the aperture nor the dynode multiplier structure. Electrical charge on the photosensitive surface of the vidicon changes with the level of incident light. The amount of charge that is replaced by a scanning electron beam is measured, along with the position of the beam, to give information about the position and intensity of the image.

In summary, there are a variety of different devices that can determine the position of a laser beam as it strikes a detector.

B. Laser Tooling

Let us consider the problem of positioning objects along a straight line. An obvious method is to project a beam of light and measure the deviation of the objects from the center of the beam. This procedure offers many advantages, especially if the alignment is to be carried out over great distances. By contrast, mechanical methods are difficult and cumbersome. Laser alignment systems have been used to align large structures and large machines to greater accuracy and at lower cost than is possible with mechanical methods.

The practice of optical alignment predates the invention of the laser and has come to be called optical tooling. The term "optical tooling" originated in the aircraft industry. Optical methods for alignment of components in an aircraft (such as wing or fuselage sections) have been in use for many years. Optical tooling at first employed instruments such as autocollimators to measure rotation of mirror targets and precision alignment telescopes, capable of measuring the displacement of objects from the optical axis of the telescope.

The development of the laser has led to significant advances in optical tooling. Low-power lasers provide considerably higher radiance for tooling applications than conventional light sources. They are easily visible in ambient lighting conditions even at distances of hundreds of feet from the laser. Lasers have been specifically designed for tooling applications, and these are sometimes called "tooling lasers." Laser tooling involves the use of the unidirectional beam for such tasks as determining displacement of objects from a line, determining angular alignment, establishing planes, performing leveling, and establishing right angles. Both helium–neon and semiconductor diode lasers have been used for this application.

The accuracy of laser tooling is limited by the angular divergence of the beam. A typical small helium–neon laser has a beam divergence angle around 10^{-3} rad. To achieve a measurement accuracy of 1 mm at a distance of 10 m, an angular spread around 0.1 mrad is required, about equal to the diffraction limit of a 1 cm aperture. The helium–neon laser beam may easily be expanded and collimated to 0.1 mrad with a small telescope. Because of its high radiance, such a beam may readily be used for alignment at large distances from the laser.

Many practical alignment systems employ a laser and a centering detector that automatically determines the location of the center of the beam. Commonly, the

quadrant photodiode discussed in the last section has been used. Accuracy of the determination of the beam center by this method is of the order of 10 μin. per foot. Thus, over distances around 100 feet, alignment can be obtained to tolerances around 0.001 in.

Lasers are specially constructed for alignment applications. On a tooling laser, the laser is designed so that the beam emerges from the center of the laser to within 0.001 in. The angle at which the beam emerges is controlled to within a few arc seconds. The laser is specially designed to provide stabilization of the beam direction.

Without stabilization, the direction of the beam can drift approximately 10 sec of arc for a temperature change of 1°C. Even after a warm-up period, fluctuations of laser temperature around 1°C/hr may occur, even if the ambient temperature is constant. Such instabilities can introduce an error of 10 sec of arc into the measurement. The stabilization techniques for high-quality alignment lasers reduce the drift to less than 1 sec of arc for a temperature change of 1°C.

Table 8-1 presents some typical specifications for commercial alignment laser systems based on helium–neon lasers. Some models include amplitude modulation of the beam and an electronic filter in the readout to reject ambient light. We note that the power is low enough so that the applications are in compliance with safety standards. Collimation of the beam leads to a beam diameter around 10 mm, which further alleviates safety problems.

The helium–neon laser was used for early generations of tooling laser systems and continues to be used. Models of commercial systems using visible semiconductor lasers with wavelengths in the 630–650 nm range are becoming common. Compared with He–Ne laser systems, diode-laser-based systems have somewhat reduced performance, but they offer the advantage of smaller and lighter packages.

Laser tooling offers several advantages over conventional tooling. Laser tooling requires only one person to set up and operate the equipment. The measurement is made by reading linear displacement directly from a meter. The readings compare well for different operators. This is in contrast to conventional optical tooling, wherein a single operator may be able to obtain reproducible readings, but different operators may arrive at different results.

Table 8-1 **Typical Specifications of Commercial Alignment Systems**

Laser power	2 mW
Long-term power drift	<5%
Beam diameter	10 mm
Beam centering on instrument	±0.001 in.
Beam alignment to instrument axis	±4 arc sec
Beam wander (after warm-up)	<20 μrad
Alignment stability (after warm-up)	±0.005 in./hr
Readout accuracy	0.001 in.
Operating range	100 m

The limitation on the use of laser alignment involves stability of the atmosphere. Beam wandering in a turbulent atmosphere will reduce the accuracy. The index of refraction of air fluctuates randomly because of temperature changes and turbulence. These inhomogeneities cause random movement of a laser beam that propagates for a distance through the atmosphere. Figure 8-2 shows some experimental results on the wandering of a laser beam caused by turbulence. These results are applicable to a propagation distance of 100 m along a path 30 cm above the ground. The measurements were taken on a windy evening at a temperature of 68°F. Beam excursions of a few centimeters were observed even for fairly short times. As the beam propagation distance increases, the amplitude of the excursions will also increase. It is apparent that beam wander of the magnitude shown in the figure will greatly reduce accuracy of an alignment application.

One simple technique to reduce beam wander, which can be employed in enclosed areas, involves an axial flow of air along the direction of alignment. A 20 in. window fan blowing along the direction of the laser beam will reduce the noise due to air turbulence.

A more sophisticated method to compensate for atmospheric turbulence involves use of centering detectors that integrate over a period of time. The distribution of light energy averaged over a suitable time interval is symmetric about the beam center. With the integrating feature, measurements of the beam center can be obtained with great accuracy, even in the presence of turbulence. Devices have been constructed that allow alignment of laser beams in a turbulent atmosphere to an accuracy of 0.2 arc sec over distances of a few kilometers [1].

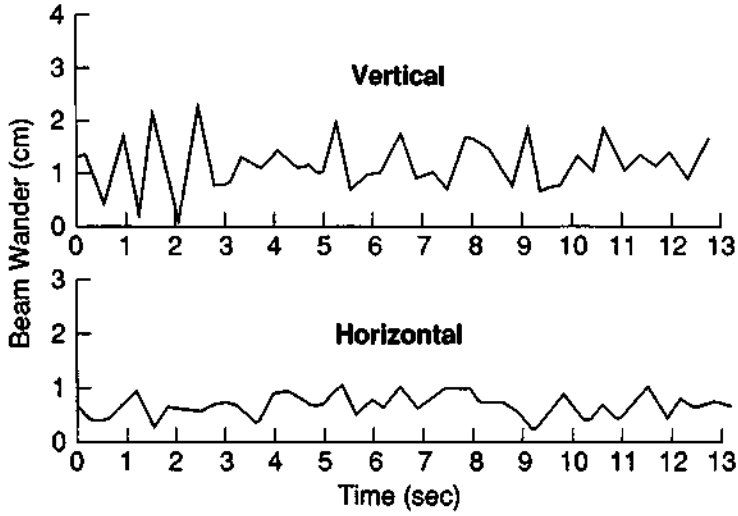


Figure 8-2 Wandering of a laser beam after traversal of a 100 m path, 30 cm above the ground, on a windy evening with a temperature of 68°F. (From A. Chrzanowski, *XIII Intern. Congr. Surveyors*, Wiesbaden, September 1–10, (1971).)

Some results showing the capabilities of angular measurements using an integrating detector at ranges out to 1500 m are shown in Table 8-2. The measurements were made under field conditions with low temperatures and high winds. The integration time was 5 sec. The movement of the laser spot due to beam wander was somewhat erratic, but generally the amplitude of the movement increased with increasing distance. The angular measurements were accurate to around 0.2 arc sec, a much better value than might be expected under the circumstances.

Another fact that must be considered for alignment over long paths is the change of air pressure (and hence of index of refraction) with height. This effect can cause light traversing a nominally horizontal path to bend downward. For typical conditions of air pressure and temperature, this effect can lead to a downward displacement of the beam by around 100 cm over a path of 10 km [2].

Let us consider the accuracy that is attainable with laser alignment. The values that have been quoted vary from one observer to another. Part of the problem is that there is no other method to check the accuracy of laser alignment. The laser provides the best alignment over large distances, and it is difficult to check independently. In a practical sense, for alignment in a shop, where doors may be opened and people are moving about, the best that seems reasonable is about 0.0002 in. over distances of 12–15 feet. The accuracy decreases as distance increases. For distances of 20 feet, perhaps 0.0005 in. is reasonable; at 50 feet, 0.001 in.; and perhaps 0.002 in. at distances of 100–200 feet. These values represent reasonable values that should be attainable in shop conditions.

Commercial versions of laser alignment systems are available and are being employed in practical applications in a number of industries, including the aircraft industry, the machine tool industry, and the construction industry. One may measure vertical and horizontal displacement of a component from the straight line reference established by a low-power laser. The line or bright spot produced by the laser beam is used to check location and orientation of components and to guide automated equipment. As specific examples, we mention the following uses:

Table 8-2 **Integrating Centering Detector Measurements**[a]

Distance (m)	Movement of laser spot (mm)	Standard deviation of angular measurement (arc sec)
100	2	0.28
200	5	0.37
300	5	0.29
400	3	0.20
500	3	0.16
600	10	0.21
800	40	0.09
1000	50	0.16
1500	40	0.20

[a]From A. Chrzanowski, *XIII Intern. Congr. Surveyors*, Wiesbaden, September 1–10, 1971.

1. Laser tooling has been used for the alignment of assembly jigs for construction of large jet aircraft [3]. For alignment of jumbo jet aircraft fuselages, lasers are superior to other methods, such as use of telescopes. Tooling holes are machined in the sections of the structure. The laser and centering detector are used to line up the sections, one by one, over distances of the order of 100 feet to an accuracy that was not previously possible. Alignment is directly read from the detector, so that operator judgment is not required. The straight line defined by the beam is available for tapping at any distance from the source. It has no sag, as any wire inevitably does. According to the users, the accuracy of alignment was ten times better than that of conventional optical systems and provided tolerances within ±0.002 in. at a distance of 200 feet.
2. The straightness of machine motion may be checked. Figure 8-3 represents measurements of the straightness of machine travel versus vertical and horizontal travel for a boring mill with a bed length of 22.5 feet. For each component of displacement, the two curves represent data obtained at two different times, three months apart. The data show that the results are reproducible to within 0.0005 in. Checking such motions with conventional techniques is extremely difficult.

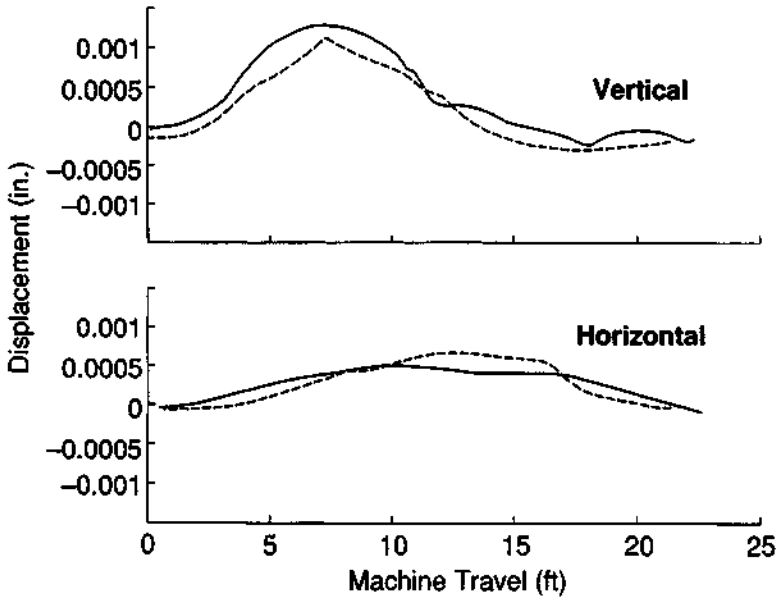


Figure 8-3 Straightness of travel of boring mill. Horizontal and vertical displacements are shown separately over a runout distance of 22.5 feet. The solid and dashed curves represent two different sets of data, taken three months apart. (From C. E. Enderby, *A Practical Laser Tooling System*, Electro Optics Assoc., Palo Alto, CA.)

Accessories are available that expand the usefulness of laser alignment systems. One device is called an optical square. This accessory, illustrated in Figure 8-4, uses a pentaprism that is rotated about the axis of the incoming beam. The pentaprism, shown in Figure 8-5, is a five-sided prism that has the property of deflecting the beam accurately by 90° from its original direction. The beam makes two reflections off aluminized surfaces and emerges in a direction perpendicular to its original direction. This establishes a right angle. The pentaprism also does not reverse or invert an image.

The pentaprism deviates the beam by 90° regardless of its orientation around the beam direction. The angular tolerance of inexpensive pentaprisms is approximately 5 arc sec. This means that for any orientation of the prism, the beam will always be deviated through the same angle, namely 90° ± 5 sec of arc. This makes the prism desirable for applications requiring right angles, without the necessity for precise orientation of the prism. Mechanical vibrations or wobble of the mechanically moving parts does not affect the accuracy of the right angle deflection.

A plane is established by rotation of the pentaprism about the direction of the incoming laser beam. The 90° reflected beam then defines a plane at right angles to the original beam. This is a simple method of defining a plane reference. Objects may easily be aligned in this plane. One may align structures, such as bulkheads or rollers, accurately perpendicular to the original laser beam.

Lasers may be used to establish a vertical direction by means of the apparatus shown in Figure 8-6. This arangement is called an optical plummet. The front surface mirror is adjusted to be perpendicular to the axis of the autocollimator telescope. Then, the laser is adjusted until the reflection of the laser beam and the reflection from the pool of mercury both coincide with the center of the reticle of the telescope. This method allows establishment of a vertical direction to an accuracy of 0.5 arc sec. The alignment may be made in a much shorter time than is possible with other techniques.

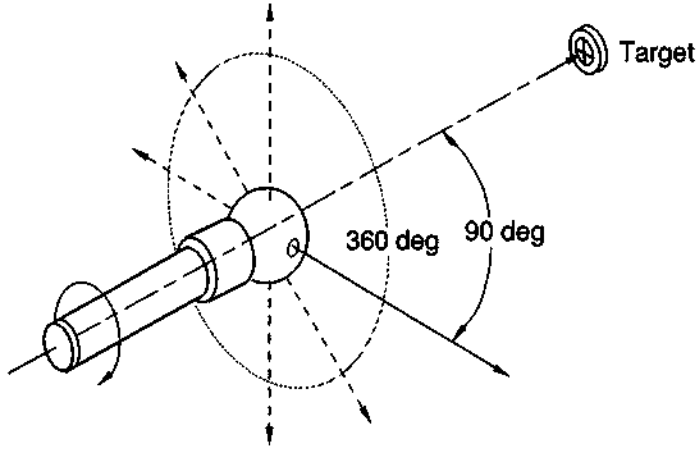


Figure 8-4 Optical square accessory.

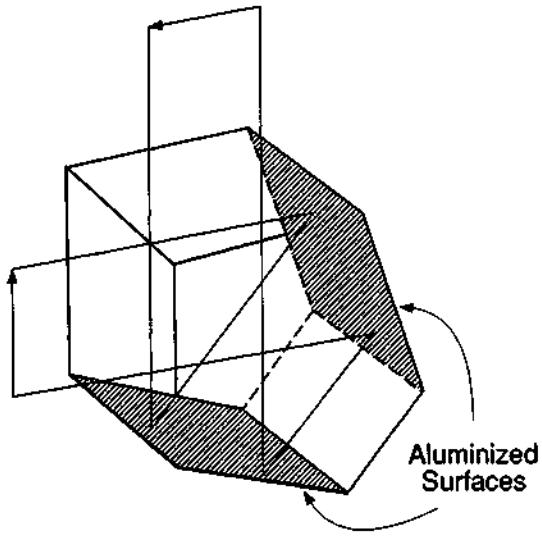


Figure 8-5 Drawing of pentaprism.

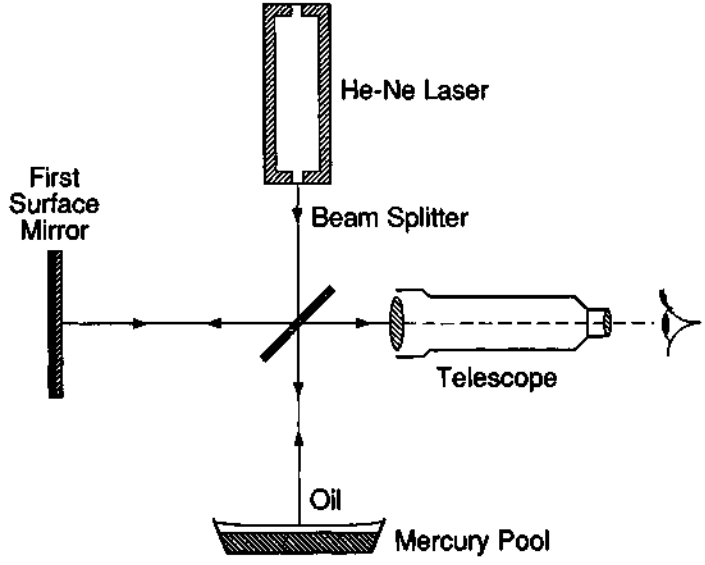


Figure 8-6 An optical plummet. (From A. Chrzanowski, F. Ahmad, and B. Kurz, *Appl. Opt.* **11**, 319 (1972).)

Autocollimating laser systems have been constructed with a centering detector built into the laser head itself. Such a device is capable of fast, accurate angular alignment when used with a conventional autocollimating target mirror.

The laser projects a collimated beam to the target mirror. If the mirror is accurately aligned, the reflected beam will reenter the front aperture of the transmitter, and the beam will return directly superimposed on the original beam. A centering detector will indicate a null.

Misalignment of the target mirror causes angular deviation of the return beam. The centering detector readout will indicate the relative angular position of the target mirror with respect to the outgoing laser beam. The centering detector can sense this offset in two axes simultaneously and display the angular deviation on two meters. Commercial models of this instrument are available, with a quoted resolution of 0.2 arc sec and an operating range of hundreds of feet.

Such an autocollimating laser may be used for measurement of angular deviation and for angular alignment of components and structures. Angular displacement may be read directly from the output of the centering detector.

As an example of the use of this device, straightness of runout of machine tools has been checked by mounting a reflector on the moving part, and monitoring its angular changes during the tool motion.

C. Angle Tracking

The receiver for an angle tracking system, in its simplest form, consists of a position-sensitive detector and an objective lens with a known focal length, which collects light and focuses it on the detector. A laser source illuminates a target. Light reflected by the target is imaged on the detector. Angular information about the target is derived from the known position sensitivity of the detector as indicated in Figure 8-7. From geometrical optics, we derive the angle α between the incoming rays and the optical axis common to the objective and the detector (the "boresight") as

$$y = f \tan \alpha \tag{8.4}$$

where y is the lateral displacement of the image from the centerline of the detector and f is the focal length of the lens. For small values of α, less than 5°, this may be approximated by

$$y = f\alpha \tag{8.5}$$

where α is in radian measure.

The diameter of the image of the target depends on the distance of the target from the receiver. If the target is at a distance much larger than any of the other dimensions, the image will essentially be a point, and the measurement of y will be more accurate than for shorter target distances.

We note that a system using a lens with a relatively long focal length produces greater displacement in the focal plane for a given angular orientation of the target.

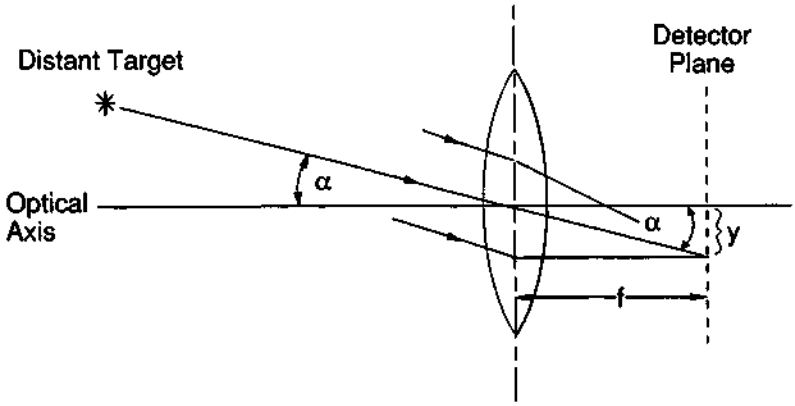


Figure 8-7 Angle tracking receiver. The lens forms an image of a distant target at angle α from the optical axis on a position-sensitive detector.

Thus, objectives with long focal length are more sensitive for detecting angular orientation.

Angle tracking systems have been used frequently for tracking of satellites and missiles. As an example, a continuous laser tracking system has been used at White Sands Missile Range for tracking of missiles from launch. A 5-W argon laser was used. The beam was scanned by a piezoelectric deflection system. The beam tracked the nose cone of the missile and was kept centered on the nose cone. The system determined both range and angular orientation of the missile. The angular tracking was performed as just described with an image dissector tube. Range information was obtained by beam modulation telemetry, a method to be discussed in Chapter 10. The azimuth and elevation angles were determined to a precision of 0.3 arc min, and the range to a precision of 0.1 m. This example shows the versatility of laser angle tracking systems.

A similar arrangement is used in industrial applications to determine the location of surfaces and the dimensions of products in a triangulation measurement. This will be described in Chapter 11.

A related angular measurement involves shaft-to-shaft alignment, in which two rotating shafts are to be aligned so that the directions of the shaft centerlines are parallel. This involves measurement of four angles, two perpendicular components for the angle each shaft centerline makes with the line between the two shafts. In one manifestation, a unit containing a semiconductor diode laser operating at 670 nm and a position-sensitive detector is mounted on each shaft using a V-block assembly. The distance between the lasers is required as an input parameter. The devices measure the four angles and specify the amount of horizontal and vertical adjustment that is required to bring the shafts into alignment. The system is quoted as having an accuracy of 0.001°.

D. Lasers in Construction

Precision optical equipment has long been used by contractors and civil engineers. Optical instruments such as transits and theodolites are essential tools of the construction industry. These instruments have been supplemented by laser-based equipment that is well accepted as a versatile labor-saving construction tool. Lasers are used to align components in building and construction jobs, to determine line and grade in excavations, tunnels, and piping, to position floating marine equipment, to provide accurate alignment in the building of bridges, sewer pipes, and waterways, to establish accurate blasting patterns, and to guide automated equipment like boring machines.

Systems have been developed for determining and controlling line and grade in sewer pipes, for example. These have reached the status of common field use.

The laser most used in construction has been the helium–neon laser. Specially designed lasers are built ruggedly for use in construction applications. Construction lasers incorporate rugged, long-life laser tubes, dust-proof housings, and moisture-proof seals. These lasers are durable, lightweight, battery-operated, and otherwise compatible with use in the field.

The applications for laser construction systems are too numerous to compile a complete list. Rather, we shall present some representative examples.

1. STRAIGHT LINE PROJECTION

One common and versatile construction laser system is one that projects a straight line. The ability of the laser to generate an accurate straight line has gained wide acceptance in the construction industry. Laser alignment is used for aligning pipe, railroad tracks, utility poles, fence posts, highway median paint stripes, and bridge pilings and in other applications where a straight line is required over long distances.

In particular, the laser has become a very common tool for aligning pipe. A number of companies have developed special laser systems for this application. Helium–neon lasers have generally been used. These systems have specially designed features, such as settings for determining the percentage of grade, digital readouts for indicating the grade, and automatic self-leveling controls.

The laser is supported at the entrance of a section of pipe with the beam projected into the pipe. The slope or grade is set by adjusting the vertical direction of the beam. A visible spot is seen on a translucent target placed at the far end of the section of pipe. Special fittings are available that are inserted into the pipe to hold the target. The pipe is moved until the laser spot is positioned at the center of the target.

Under some conditions, temperature differences will exist inside the pipe. Such a temperature difference could be caused by sunlight heating the top of the pipe while contact with the ground cools the bottom. The temperature difference causes the index of refraction of the air to change, leading to a bending of the light beam.

This would lead to inaccurate alignment of the pipe. The inaccuracy can be eliminated by using a blower to circulate the air in the pipe in order to equalize the air temperature.

Complete laser alignment systems designed for field use in pipeline construction are available. These eliminate the need for batterboards (pairs of horizontal boards nailed to posts to indicate a desired level), line and grade strings, and the associated optical equipment that previously had been used for this type of work. As a result, the setup costs are reduced and construction is speeded. Pipe laying accuracy is also increased. The laser systems produce more accurate references than previous systems, so that pipe can be laid to a more precise line and grade. Such systems have practical use in sewer installation, where a precise grade must be maintained.

Lasers have also been used to project straight lines for applications like tunneling. The operator of the tunneling equipment keeps the projected laser beam on a target mounted on the equipment, ensuring that the equipment continues to move in a straight line. One well-known application of this type was the construction of the Bay Area Rapid Transit System in California. A helium–neon laser was used to guide the tunneling machinery. The resulting tunnels were straight within approximately 1 in. over a 1.5 mile span.

Another example involved alignment of the two-mile-long Stanford linear accelerator [4]. This application employed diffraction gratings to increase the precision of the optical alignment. The diffraction gratings led to a more compact intensity pattern at the measurement point. A laser-illuminated Fresnel zone plate was used to align to an accuracy of 0.5 mm at a distance of 3 km. Such tolerances are beyond the capabilities of conventional methods.

One may also align objects along the straight line defined by the laser beam by using a centering detector mounted on the components. This is done in the same manner as discussed earlier in this chapter.

In overwater construction, operators can line up a barge with reference to the beam from a laser mounted on the shore. This can lead to savings because the operator can make quick intermediate alignment control checks.

A number of straight line projection laser systems have been designed. Some incorporate a sighting telescope boresighted with the laser beam. The laser beam and the telescope can be used independently to determine identical points.

2. *PLANE-OF-LIGHT SYSTEMS*

When a laser beam is reflected by a rotating prism or mirror, a plane of light is produced. If the angle of the laser beam is adjusted, the plane of light can be projected at any angle, from horizontal to vertical. We have already discussed the use of a rotating pentaprism to define a plane of light, but other types of prisms and mirrors have been used to scan the beam. In some cases, the element scans less than 360°, and only a partial plane is defined.

Floors, ceilings, and elevation benchmarks can be positioned using a rotating laser beam in a horizontal plane. Only one laser position is needed for a given job.

Several workers can perform individual tasks at different locations around the same laser reference beam.

In particular, installation of ceiling tile using a plane of light generated by a helium–neon laser and a rotating pentaprism has become common. The laser beam is accurately aligned with the vertical so that the plane of light is horizontal. Workers move the supporting grid for the tile up or down until it is in line with the plane of light, and then they tie the grid at that position. This method may save 25–50 percent of the labor required to install ceiling tile, and it is more accurate than conventional methods based on strings and wires.

A vertical plane of light can be produced by orienting the laser beam horizontally and directing it to a rotating pentaprism. A vertical plane of light can serve as a reference for applications like aligning tilt-up wall sections.

The helium–neon laser has been the type most often employed for applications of this sort, but visible semiconductor diode lasers are now also being used.

3. *HEAVY EQUIPMENT CONTROL*

Equipment guidance systems utilize a rotating laser beam, a sensor mounted on the machine, and a readout or automatic control unit. As an example, consider control of road building equipment, such as graders. One can set exact grade levels. A number of different types of systems have been devised. For simplicity, we shall describe one specific example. In this example, the system uses a tripod-mounted laser with a rotating prism, a sensor that mounts on the equipment, and a readout instrument on the operator's control panel. The laser and the rotating prism can project a level plane. Or if the work calls for a slope or grade, the plane can be tilted at the desired angle.

The sensor is mounted on a mast high enough above the other portions of the machinery that the beam can be detected, no matter what the relative orientation of the machinery and the laser. The sensor detects whether the position of the plane of light striking it is at the right level, or high or low. Lights and audible signals inform the operator of the condition. The operator holds the reading on center by adjusting the blade level or depth of cut. This system allows the blade level to be held within ±0.25 in. over distances up to 1500 feet from the laser.

This particular system involves operator control to complete the feedback loop. More sophisticated systems with servo control can provide automatic adjustment and control.

A laser system of this type offers reduced time and labor compared with the use of a stretched wire grading reference. It is much easier to project a laser beam across a field as an elevation reference than it is to stretch a wire or set up grade targets.

In summary, laser equipment has gained considerable acceptance in many highly competitive construction applications.

References

[1] A. Chrzanowski, F. Ahmad, and B. Kurz, *Appl. Opt.* **11**, 319 (1972).
[2] J. C. Owens, Laser Applications in Metrology and Geodesy, in *Laser Applications 1* (M. Ross, ed.), Academic Press, New York, 1971.
[3] S. Minkowitz, *Electro-Opt. Systems Design*, p. 22 (April 1970).
[4] W. B. Hermannsfeldt, *et al.*, *Appl. Opt.* **7**, 995 (1968).

Selected Additional References

J. Cornillault, Using Gas Lasers in Road Works, *Appl. Opt.* **11**, 327 (1972).
P. W. Harrison, Alignment Techniques, in *Optical Transducers and Techniques in Engineering Measurement*, (A. R. Luxmoore, ed.), Applied Science Publishers, London, 1983.
J. T. Luxon and D. E. Parker, eds., *Industrial Lasers and Their Applications*, 2nd edition, Prentice-Hall, Englewood Cliffs, NJ, 1992, Chapter 9.

Chapter 9 | Principles Used in Measurement

Some of the most important applications of lasers involve their use for measurement of a wide variety of different quantities: distance, product dimension, angular rotation rate, and so forth. Optical methods of measurement have been employed for many years. Previously, they had been limited by the available light sources, which traditionally lacked sufficient coherence or sufficient brightness to perform all desired measurement tasks well. The development of the laser has removed these limitations. The availability of lasers has led to improvement of conventional optical measurement techniques and the introduction of many effective new techniques. Laser-based measurements have been particularly useful for remote, noncontact measurement.

Laser-based measurement provides a versatile tool capable of being used for a variety of different measurements. In this chapter, we will describe principles that are common to many different measurement applications. In the following chapters, we shall discuss specific applications, including measurement of distance, velocity, fiber or wire diameter, angular rotation rate, and material thickness.

The principles described in this chapter have been utilized in many different situations. The same physical principles recur in various applications. It will assist understanding these applications if we first review some of the physical phenomena that are often employed. Some of the most common measurements employ interference phenomena, which arise when two beams of light are combined. Our discussion will emphasize interference.

We first describe the Michelson interferometer, which produces an interference pattern between two beams of light that traverse different paths. We then describe production of an oscillating beat note, which involves interference of two light beams of slightly different frequency. We then discuss the production of a frequency shift using the Doppler effect, and the role of the coherence of laser light in the ability to produce usable interference patterns.

A. The Michelson Interferometer

The Michelson interferometer consists of two mirrors, M_1 and M_2, arranged as shown in Figure 9-1, with a beamsplitter inclined at 45° to the mirrors. The collimated beam of laser light is incident on the beamsplitter, and it is divided into two beams when it strikes the partially reflecting surface on the beamsplitter. Part of the light follows the path to mirror M_1, is reflected by M_1, and retraces the path to the beamsplitter. Some of this light will pass through the semitransparent beamsplitter. Another light ray originally passes through the beamsplitter to mirror M_2, is reflected back to the beamsplitter, and is partially reflected by the surface of the beamsplitter, so as to be superimposed on the first ray. These two beams contain only part of the original light, the rest having been lost by reflection or transmission

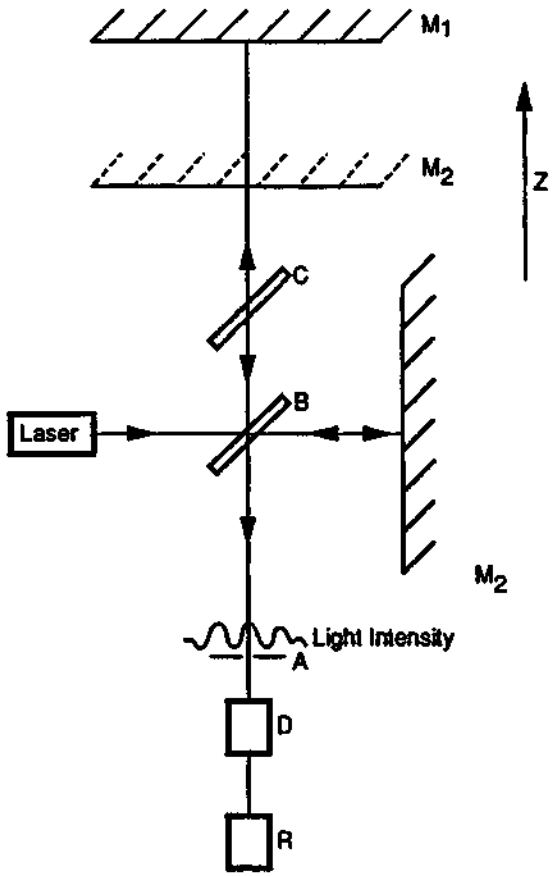


Figure 9-1 Schematic diagram of Michelson interferometer. In this diagram, M_1 and M_2 are mirrors, M_2' is the image of mirror M_2 observed when looking in the z direction, B is a beamsplitter, C a compensator, A an aperture, D a detector, and R a recorder.

at the beamsplitter and returned back toward the laser. A compensating plate, which has the same thickness and is made of the same glass as the beamsplitter, is sometimes placed in the path of the first ray. The first ray traverses the beamsplitter only once, whereas the second ray traverses the beamsplitter three times. Insertion of the compensating plate provides for equal paths in glass.

The mirror M_2 has an image parallel to M_1 because of reflection at the semireflecting surface of the beamsplitter. This image is indicated by the dotted surface M_2'. The Michelson interferometer produces interference that is identical to the interference produced by the wedge of air between the mirror M_1 and the image M_2'. An observer views the light emerging from the beamsplitter containing the superimposed reflections from M_1 and M_2'. These reflections are seen by the observer in the same direction. The path difference between the beams is $2S \cos \theta$, where S is the distance between M_1 and M_2' and θ is the angle between the axis of the system and the direction at which the rays are observed. The phase difference δ between the two beams is

$$\delta = 2\pi \frac{2S \cos \theta}{\lambda} \tag{9.1}$$

The formation of fringes is a result of the wave nature of light. A light wave moving in the negative z direction may be described by the function

$$E(z, t) = E_0 \exp i(kz + \omega t + \phi) \tag{9.2}$$

where E_0 represents the amplitude of the wave, ω the angular frequency, and $k = 2\pi/\lambda$, with λ the wavelength. The function ϕ represents the phase of the wave, and it is generally a function of position and time. For monochromatic and coherent laser light, we may consider ϕ as a constant.

The two beams of light from M_1 and M_2' combine to yield an intensity I given by

$$I = |E_1 \exp i(kz + \omega t + \phi_1) + E_2 \exp i(kz + \omega t + \phi_2)|^2 \tag{9.3}$$

where E_1 and E_2 are the amplitudes of the two beams and ϕ_1 and ϕ_2 are their phases.

Equation (9.3) represents the so-called "square law." When two light beams are combined, the resulting intensity is given by summing the amplitudes of the two beams and then taking the square of the absolute value of the sum. Detectors of light, including phototubes, photodiodes, photographic film, and the human eye, all respond to the intensity of the light as given by the square of the absolute value of the amplitude. This is in contrast to microwaves, where linear detectors, which respond directly to the amplitude, are available. If we carry out the operations indicated by Equation (9.3), we obtain

$$\begin{aligned} I &= [E_1 \cos(kz + \omega t + \phi_1) + E_2 \cos(kz + \omega t + \phi_2)]^2 \\ &\quad + [E_1 \sin(kz + \omega t + \phi_1) + E_2 \sin(kz + \omega t + \phi_2)]^2 \\ &= E_1^2 + E_2^2 + 2E_1E_2 \cos(\phi_1 - \phi_2) \\ &= E_1^2 + E_2^2 + 2E_1E_2 \cos \delta \end{aligned} \tag{9.4}$$

where δ is given by Equation (9.1). This result in not simply an exercise in trigonometry; it expresses the pattern of light intensity really seen by a detector. The result expressed by Equation (9.4) corresponds to a general background intensity $E_1^2 + E_2^2$ modulated by fringes expressed by the cosine term. If the amplitudes are adjusted to be equal, then

$$I = 2E_1^2(1 + \cos\delta) \tag{9.5}$$

and the light intensity is zero when

$$\delta = (2M + 1)\pi \tag{9.6}$$

where M is an integer. This corresponds to formation of a dark fringe. The light intensity is a maximum at

$$\delta = 2M\pi \tag{9.7}$$

This corresponds to formation of a bright fringe in the interference pattern. The fringe will be circular, because a given value of phase difference corresponds to a circle about the line from the eye normal to the mirror. At the center of the fringe pattern, where $\theta = 0$, we have $2S = M\lambda$ (with M an integer) as the condition for a bright spot.

Consider the condition where M_1 is more distant from the beamsplitter than M_2' is. A pattern of circular fringes is seen. The fringes are packed closer together as their radii increase. As M_1 is moved away from M_2', the fringes move out from the center and the radii increase. A new fringe appears at the center of the pattern for each movement of the mirror through a distance $\lambda/2$. In other words: as the path difference between M_1 and M_2' changes by half the wavelength, the light intensity varies from maximum to minimum and back to maximum. The variation along the axis ($\theta = 0$) can be detected by insertion of an aperture as shown in Figure 9-1, with a photodetector behind the aperture. Counting the number of cycles of oscillation of light intensity then gives the distance the mirror M_1 has moved, with each full oscillation of light intensity corresponding to a motion of one-half wavelength.

B. Beat Production (Heterodyne)

Another type of interference effect that is often employed in measurement is the phenomenon of a beat note. If two light beams of slightly different frequency interfere, the interference fringes will not remain fixed in time, but will be in motion, At a given position in space, the motion of the interference pattern will produce a varying intensity.

The phenomenon is analogous to the familiar case in acoustics in which two tuning forks of slightly different frequency are struck simultaneously. Interference of the two sound waves produces a sound of oscillating amplitude. This phenomenon is referred to as beating, and the frequency of the variation of the sound is the beat frequency. As utilized in optics, this is sometimes called optical heterodyne or photomixing.

We shall treat this phenomenon similarly to the discussion centering on Equations (9.3) to (9.5). As before, we represent the light wave traveling in the negative z direction by the functional form given in Equation (9.2). We note again that photodetectors respond to the square of the amplitude of the optical field. The intensity I of the two combined beams is

$$\begin{aligned} I &= |E_1 \exp i(k_1 z + \omega_1 t + \phi_1) + E_2 \exp i(k_2 z + \omega_2 t + \phi_2)|^2 \\ &= [E_1 \cos(k_1 z + \omega_1 t + \phi_1) + E_2 \cos(k_2 z + \omega_2 t + \phi_2)]^2 \\ &\quad + [E_1 \sin(k_1 z + \omega_1 t + \phi_1) + E_2 \sin(k_2 z + \omega_2 t + \phi_2)]^2 \\ &= E_1^2 + E_2^2 + 2E_1 E_2 \cos[(k_1 - k_2)z + (\omega_1 - \omega_2)t + \phi_1 - \phi_2] \\ &= E_1^2 + E_2^2 + 2E_1 E_2 \cos[(k_1 - k_2)z + (\omega_1 - \omega_2)t + \delta] \end{aligned} \tag{9.8}$$

where δ is again the difference in phase between the two beams, and where now we consider the beams to have different angular frequencies, ω_1 and ω_2.

At a fixed position, corresponding to given values of z and δ, the light intensity viewed by a detector will contain a time-varying component with angular frequency $\omega_1 - \omega_2$, that is, the difference in angular frequency between the beams. This oscillating signal is referred to as a beat note.

Derivation of Equation (9.8) in this way shows how production of the beat frequency is related to production of fringes as expressed by Equation (9.4). Indeed, this is a generalization of the earlier derivation, and it contains the result that if the two beams have different frequency, the fringe pattern will be in motion. For a constant position on the z axis, the bright fringes will satisfy

$$(\omega_1 - \omega_2)t + \delta = \text{constant} \tag{9.9}$$

where δ may be considered to be a function of the viewing angle θ.

The variation in the optical signal itself, which is of the order of 10^{15} Hz, is too fast for photodetectors to follow. But if the angular frequencies ω_1 and ω_2 are reasonably close together, the difference $\omega_1 - \omega_2$ will be small enough for the detector to follow. Then, the output of the detector will contain a term oscillating at the difference frequency. The output of a photodetector viewing the two combined laser beams with slightly different frequencies does in fact contain this time-varying component.

Many measurement applications involve a shift in frequency of part of the beam from a laser. Subsequent recombination of the original beam and the frequency-shifted beam yields the beat note, the frequency of which can be measured, giving in turn a measurement of the parameter that caused the shift. This technique proves useful for a varied range of measurements.

C. The Doppler Effect

A common method for providing the frequency difference that can be employed to produce a beat note involves the Doppler effect. When light of frequency f is re-

flected from a surface moving at velocity v, there is a change in frequency because of the Doppler effect. The optical Doppler effect is analogous to the familiar acoustic frequency shift that is observed from moving sources of sound waves. The magnitude δf of the Doppler shift is

$$\frac{\Delta f}{f} = \frac{2v}{c} \tag{9.10}$$

where c is the velocity of light.

Because v is much smaller than c, the frequency shift will be a small fraction of the original frequency. For example, a velocity of 1500 cm/sec (a little more than 30 miles/hr) yields $\Delta f/f = 10^{-7}$, or $\Delta f = 10^8$ Hz. Two conclusions are apparent:

1. A well-stabilized laser must be used to be able to detect the frequency shift.
2. Frequency shifts arising from Doppler shifting of light can be in the range where measurements can easily be made.

Thus, the Doppler effect can provide a convenient source of two light beams at slightly different frequencies. The beam from a laser is split by a beamsplitter; one of the beams is Doppler shifted; and then the beams are recombined at a detector in order to detect the beat frequency arising because of the Doppler shift. Because both beams originate from the same laser, the small difference in frequency can be carefully controlled, to a degree that would not be possible if two independent lasers were used.

D. Coherence Requirements

The laser is useful for the measurements described here because the properties of the wavefront do not change erratically with time and with position; that is, the laser light is coherent. The electric field associated with the light wave may be written

$$E(\mathbf{r}, t) = A(\mathbf{r}, t) \exp i[\mathbf{k}\cdot\mathbf{r} - \omega t + \phi(\mathbf{r}, t)] \tag{9.11}$$

where A, the amplitude function, and ϕ, the phase function, are both functions of position $\mathbf{r}$ and time t, and $\mathbf{k}$ is a vector with magnitude $2\pi/\lambda$ and the same direction as that of the light propagation. Consider first the time variation. If A and ϕ are slowly varying compared with the function $e^{i\omega t}$, then during one period of the light wave, the functions A and ϕ can change only slightly. In our previous analysis, we assumed that for the time periods of interest, A and ϕ are constant. However, in reality, these functions will fluctuate to some extent and reduce the coherence of the system.

The effect of the variations in the phase function ϕ will be to decrease the visibility of the interference fringes. If the time of measurement is long compared with the characteristic time over which ϕ varies, the cosine term in the last line of Equation (9.4) will fluctuate. The average light intensity in the interference pattern will reach some common average value, and the fringes will disappear.

To preserve the visibility of the fringes, the coherence time of the laser must be long enough for the two beams traversing the different arms of the interferometer to remain mutually coherent when they are recombined. In other words, the coherence length of the laser must be longer than the path difference in the two arms. This leads to the condition

$$S < \Delta L = c\,\Delta t \approx \frac{c}{\Delta \nu} \tag{9.12}$$

where ΔL is the coherence length of the laser, Δt the coherence time, $\Delta \nu$ the spectral width of the laser line, and S the path difference between the two arms of the interferometer.

Some typical values for coherence length are given in Table 9-1. The high-pressure mercury arc has too short a coherence length to make useful interferograms. Some low-pressure arc lamps have longer coherence length, but their intensity is much reduced and the interference patterns are not bright. The fluorescent linewidth of neon (which defines the maximum linewidth for a multimode helium–neon laser) would yield a coherence length of around 20 cm. Thus, one could use a multimode helium–neon laser to make interferograms provided that the maximum path difference is no more than a centimeter or two. A single-mode helium–neon laser has coherence length great enough to make measurements even when the path difference is significant.

Similar remarks apply to beat production. The coherence time must be long compared with the period of the beat. This leads to the requirement

$$\frac{1}{\Delta \nu} > \frac{1}{f_1 - f_2} = \frac{1}{\Delta f} \tag{9.13}$$

where Δf is the beat frequency. For shifts that occur by virtue of the Doppler effect, the requirement is

$$\frac{1}{\Delta \nu} > \frac{1}{2vf/c} \tag{9.14}$$

or

$$\Delta \nu < \frac{2vf}{c} \tag{9.15}$$

Table 9-1 **Coherence Properties of Light Sources**

Source	Linewidth $\Delta \nu$ (Hz)	Coherence length ΔL (cm)
High-pressure mercury arc	5×10^{12}	0.006
Low-pressure arc lamp	10^9	30
Multimode He–Ne laser	Up to 1.5×10^9	20
Single-mode He–Ne laser	10^6	3×10^4

This sets a lower limit to the velocity that can be measured by a frequency shift using the Doppler effect, with a laser having a defined spectral width.

The spatial coherence properties of the laser are also important. Spatial coherence is a measure of how the phase function varies between different positions on the wavefront at one instant of time. Conventional nonlaser light sources have poor spatial coherence. To improve the spatial coherence of nonlaser sources, the area of the source may be masked, so that only a small fraction of the source can contribute to the interference pattern. This increases the spatial coherence sufficiently so that fringes can be formed. But this reduces the amount of light available and makes the fringes faint.

The spatial coherence of a laser is related to the transverse mode structure. Any one transverse mode of a laser will be spatially coherent. However, the lowest-order TEM_{00} mode will provide the most uniform illumination. Most interferometric applications of lasers use devices operating in the TEM_{00} mode.

The high radiance of the laser is related to these considerations. All the light from the laser may be employed for forming the fringe pattern, in contrast to the situation for nonlaser sources. The laser beam is well collimated, so that all the light is easily collected. Thus, fringes formed by a laser source are easily visible under conditions where the interference pattern formed by a nonlaser source would not be visible.

Selected References

A. The Michelson Interferometer

M. Born and E. Wolf, *Principles of Optics*, 6th edition, Pergamon, Oxford, 1980, Chapter 7.

M. Francon, *Optical Interferometry*, Academic Press, New York, 1966, Chapter IV.

F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th edition, McGraw-Hill, New York, 1976, Chapter 13.

B. Beat Production (Heterodyne)

E. B. Brown, *Modern Optics*, Van Nostrand-Reinhold, Princeton, NJ, 1965, Chapter 13.

M. Ross, *Laser Receivers*, Wiley, New York, 1966, Chapter 4.

A. Yariv, *Optical Electronics*, Holt, Rinehart and Winston, New York, 1985, Chapter 11.

C. The Doppler Effect

J. C. Angus *et al.*, Motion Measurement by Laser Doppler Techniques, *Ind. Eng. Chem.* **61**, 8 (1969).

A. E. Siegman, *Lasers*, University Science Books, Mill Valley, CA, 1985, Chapter 25.

D. Coherence Requirements

R. J. Collier, C. B. Burckhardt, and L. H. Lin, *Optical Holography*, Academic Press, New York, 1971, Chapter 7.

H. M. Smith, *Principles of Holography*, Wiley (Interscience), New York, 1969, Chapter 6.

Chapter 10 | Distance Measurement and Dimensional Control

This chapter begins the description of laser applications as sensors and measurement tools. The emphasis in this chapter is on distance measurement and the associated uses for dimensional control. The next chapter will describe a variety of other laser sensing and measurement techniques.

Lasers can be applied in a variety of ways to solve practical problems in distance measurement. The most important methods that have been employed are summarized in Table 10-1, along with some of the areas in which the methods have been applied. In the table, the accuracy is specified as a fraction of the distance being measured. Interferometric systems, usually using helium–neon lasers, offer precise distance measurement over a scale of distances less than 100 m and in an indoor environment. Such devices are suitable for dimensional control of machine tools. A method that competes with interferometric distance measurement is laser Doppler displacement. In this approach, the Doppler shift of the beam reflected from a target is measured and integrated to obtain displacement. This method also is best suited for use indoors at distances no more than a few hundred feet.

For longer distances and for use in the field, other methods are more appropriate. A method often used outdoors is beam modulation telemetry. In this method, a modulated beam is transmitted to a distant object, and a reflected return signal is detected. The phase of the modulation of the return signal is compared with the phase of the outgoing signal. This method is useful for distances of hundreds of meters and is used in surveying.

For other applications, especially over longer distances and in cases where somewhat less precision may be tolerated, the round-trip transit time for a very short high-power pulse of light can be measured. This approach has most often been used for military range finding.

Table 10-1 **Laser-Based Distance Measurement**

Method	Typical laser	Range	Typical accuracy	Typical application
Interferometry	He–Ne	Meters (indoors)	Microinch (typically 1 part in 10^7)	Machine tool control, table flatness, seismic and geodetic uses, length and angle standard calibration
Laser Doppler displacement	He–Ne	To 400 feet (indoors)	Microinch (1 part in 10^7)	Machine tool control, table flatness, vibration, angular calibration
Beam modulation telemetry	He–Ne, AlGaAs	Up to kilometers	1 part in 10^6	Surveying
Pulse time of flight	Q-switched solid state	Many kilometers	1 part in 10^5	Satellite ranging, military ranging

A. Interferometric Distance Measurement

Interferometric measurement of distance with optical sources is not new. But lasers offer two important advantages over conventional light sources for interferometric measurement of distance:

- High radiance
- High temporal coherence

These properties allow precise measurements to be made over considerable distances, greater than would be possible with other light sources.

Most interferometric systems use a helium–neon laser operating at a wavelength of 632.8 nm (corresponding to a frequency around 5×10^{14} Hz). For an unstabilized laser, there would be a possible variation in frequency of 10^9 Hz, approximately the width of the fluorescent line in neon. This would mean an uncertainty of 1 part in 5×10^5 in the measurement, because the laser could operate on more than one longitudinal mode that falls within the linewidth. Thus, for interferometric applications, the laser must be stabilized so that the operating frequency falls at the center of the neon fluorescent line. A typical design for frequency stabilization involves construction of a relatively short laser so as to provide operation in a single longitudinal mode. The mirrors are rigidly mounted on a single piece of material with very low thermal expansion, such as invar. The laser is contained in a temperature-controlled environment, such as an oven, to provide additional thermal stability. Then the length of the cavity is servo controlled to the position of a small dip in power that occurs when the laser frequency lies at the center of the neon fluorescent line.

As the length of the laser operating in a single longitudinal mode varies slightly, the power output varies. When the cavity length is adjusted to the exact center of the fluorescent line, there is a slight local decrease in the laser power. This decrease is called the Lamb dip. Servo control of the cavity length using piezoelectric transducers allows the laser output to be stabilized at the frequency corresponding to the center of the fluorescent line. The frequency stability obtained in this way is approximately 1 part in 10^9 of the operating frequency.

Making the laser cavity short constrains the laser to operate in a single longitudinal mode. The spacing between longitudinal modes is $c/2D$, where c is the velocity of light and D is the length of the laser cavity. Thus, shortening the laser cavity will increase the mode separation, until the separation becomes comparable to the width of the fluorescent line of neon. Then, there can be only one longitudinal mode present within the fluorescent linewidth. This leads to operation at a single frequency. Single-mode operation is achieved at an expense of laser output power, because of the decreased volume of the active medium. Such mode-stabilized helium–neon lasers typically emit powers at levels of a fraction of one milliwatt.

We describe first an interferometric method based on the Michelson interferometer, because the basic operation of interferometric distance measuring systems can be understood easily with reference to this interferometer. Later, we will describe variations that offer better performance under conditions of atmospheric turbulence and attenuation. The Michelson interferometer has been described in Chapter 9. A schematic diagram of an interferometric distance measuring system is shown in Figure 10-1. The laser beam is split into two parts, a measurement beam and a reference beam. Each of these beams traverses one arm of the interferometer. The measurement beam travels to a mirror mounted on the moving part, the distance of which is to be measured. The mirror usually is usually a cube corner reflector. The cube corner reflector is a trihedral prism that acts as a retroreflector. The light undergoes one reflection at each face of the prism and returns in a direction almost exactly antiparallel to the direction from which it originally came (see Figure 10-2). The retroreflecting feature is not critically dependent on the exact angular alignment of the prism. Thus, tolerance to small alignment errors and to vibration is much greater than if a conventional mirror were employed.

The two beams are recombined at a beamsplitter and travel to a detector. The two combined beams form an interference pattern. The amplitude of the light at the detector depends on the phase between the reference beam and the measurement beam, which in turn depends on the difference in the optical paths traversed by the two beams. The phase difference δ is

$$\delta = 2kS\cos\theta \tag{10.1}$$

where $k = 2\pi/\lambda$ with λ the wavelength, S is the path difference, and θ is the angle between the common axes of the two beams and the direction of observation. Usually one chooses $\theta = 0$. When the moving part travels one-half wavelength of light, the total difference in optical path changes by one wavelength. This gives rise to the factor 2 in Equation (10.1). The phase of the light at the position of the detector will change by 2π; that is, the fringe pattern will go through one period. This will

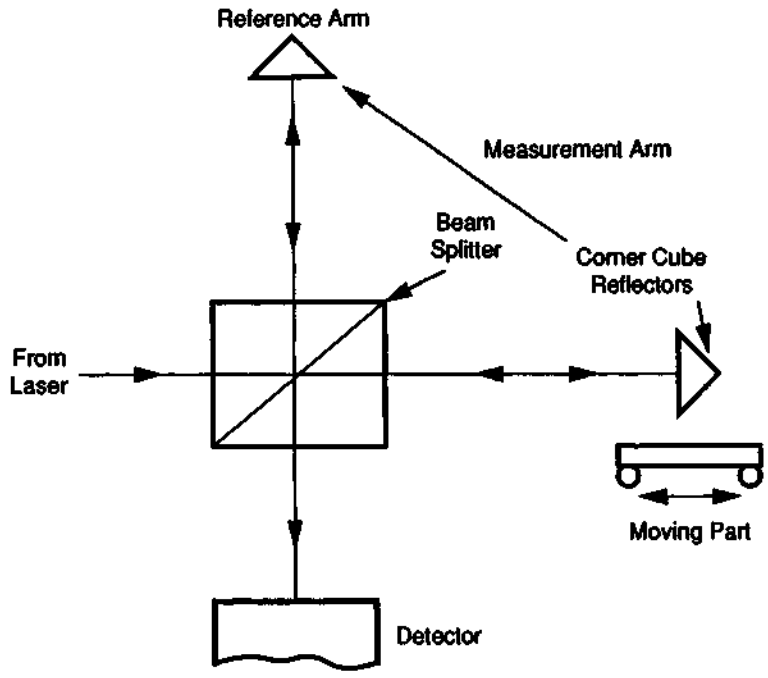


Figure 10-1 Schematic diagram of the application of Michelson interferometry to the measurement of the distance traversed by a moving part.

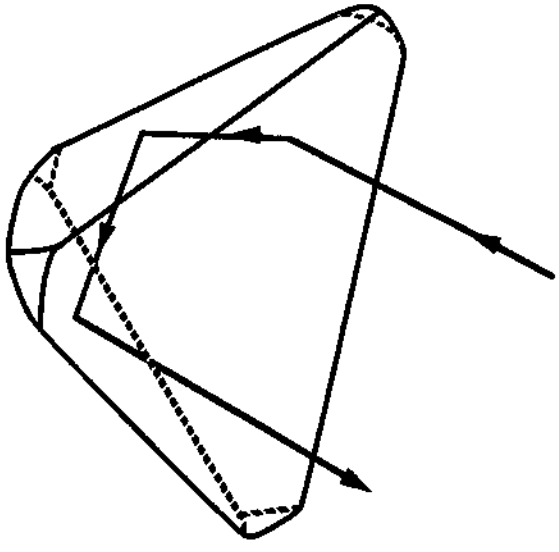


Figure 10-2 A cube corner retroreflector.

correspond to a change from light to dark to light at the position of the detector. In other words, the motion of the moving part leads to amplitude modulation of light, which is sensed by the detector. Electronic circuitry can then count the periods of amplitude modulation and calculate the distance that the part has moved.

Laser interferometers provide measurement of displacement from a starting position, rather than absolute measurement of position. The instrument reading is set to zero at the initial position of the moving part. The motion is measured relative to this zero position.

The distance measured in interferometric measurements is the optical path, which differs from the physical path by a factor equal to the index of refraction of the air. Corrections for the index of refraction often are the limiting factor in the overall accuracy of length measurement. The index of refraction n_d of dry air at 15°C and at a pressure of 760 Torr with 0.03 percent CO_2 is 1.0002765 at the helium–neon laser wavelength, which in vacuum is 0.63299138 μm. Corrections to the index of refraction n for humidity and pressure corrections are given by [1]

$$n - 1 = \frac{(n_d - 1)P}{720.775} \times \frac{1 + P(0.817 - 0.0133T) \times 10^{-6}}{1 + 0.003662T} - 5.6079 \times 10^{-8} f \qquad (10.2)$$

where f is the partial pressure of water vapor in an atmosphere of total pressure P Torr at temperature T°C. Under most conditions the effects of a varying concentration of CO_2 are small, but there may be some conditions when a high CO_2 content could cause noticeable errors.

To summarize the implications of this equation, an increase in pressure of 1 Torr increases the index of refraction by 0.36 ppm near standard conditions. A temperature increase of 1°C decreases the index of refraction by 0.96 ppm and an increase of 1 Torr in the partial pressure of water vapor decreases the index by 0.06 ppm, near standard conditions. To a reasonable approximation, these corrections may be added independently.

To convert the measured optical path length to a physical path length, these corrections must be taken into account. Systems that automatically sense the air temperature, pressure, and relative humidity, calculate the appropriate correction, and display the corrected result have been developed. In addition, the temperature of the workpiece can be measured and length measurements normalized to a reference temperature.

A schematic diagram of such a system is shown in Figure 10-3. This system uses two photodetectors to determine the direction of motion. The two detectors collect light from regions of the fringe pattern where the phase difference of the interfering beams differs by $\lambda/2$. The relative phase of the amplitude modulation viewed by the two detectors will be different depending on the direction of the motion of the remote mirror. The electronic circuitry digitizes the signal and adds a count for an increase of $\lambda/2$ in the optical path and subtracts a count for a decrease of $\lambda/2$.

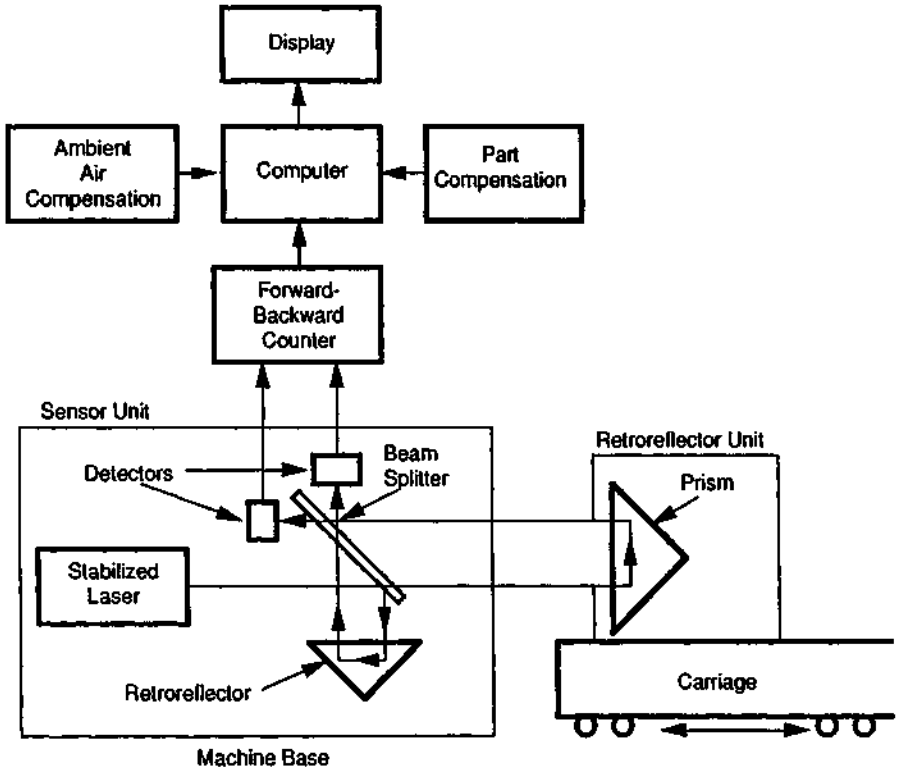


Figure 10-3 Schematic diagram of complete system for measuring motion of machine tool carriage, including compensation for ambient air parameters, part temperature, and sense of motion.

More sophisticated arrangements for determining the direction of motion have been developed. One example is shown in Figure 10-4. The laser beam passes through a quarter-wave plate and becomes circularly polarized. The beam is split by a beamsplitter to form a measurement beam and a reference beam, both circularly polarized. Each time a beam is reflected from a surface, the phase shifts by $\pi/2$, and the sense of rotation of the circular polarization is reversed. After the two beams are recombined, the net effect of all the reversals has caused the two beams to have opposite senses of rotation of the circular polarization. When two beams of opposite circular polarization are combined, the result is a beam with linear polarization. The orientation of the linear polarization will depend on the relative phase of the reference and measurement beams, and it will rotate as the retroreflector position changes. The rotation is equivalent to 180° for a target movement of one-half wavelength. Polarizers oriented at 45° to each other are in front of the detectors. The amplitude modulation of the signals from the detectors will be in quadrature. The relative phases of the amplitude-modulated signals then indicate the sense of the motion of the retroreflector.

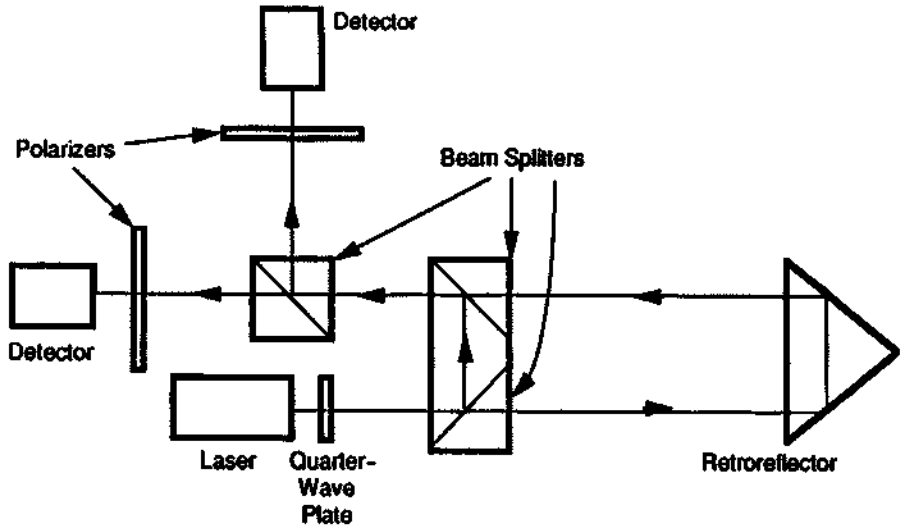


Figure 10-4 System for determining the sense of motion of a moving retroreflector.

Other variations have involved multiple passes of the measurement beam along the path to the retroreflector. This increases the sensitivity, so that in some cases one count on the counter corresponds to 1/32 wavelength.

The Michelson interferometric technique we have described here was developed early in the history of lasers, and it was used throughout the 1960s. This method of distance measurement is sensitive to variations in the air through which the beams pass. If the beam intensity changes because of air turbulence or air turbidity, improper fringe counting can result, and the measurement will be in error. These shortcomings are overcome by use of a two-frequency laser system that mixes beams of two different frequencies and measures the Doppler shift of the beam reflected from the moving retroreflector [2]. This approach, introduced in the early 1970s, greatly reduces the sensitivity of the measurement to fluctuations in the air. It quickly supplanted the simple interferometric method just described, and it is now the method of choice.

The two-frequency method is illustrated in Figure 10-5. The helium–neon laser emits light of two different frequencies f_1 and f_2, with different polarization properties. This output is achieved by applying an axial magnetic field that splits the fluorescent line of the neon into two components. The frequency-stabilized laser then emits two frequencies separated by about 2×10^6 Hz. In this system, one measures the velocity of the moving retroreflector by means of the Doppler shift, as was described in Chapter 9. Light reflected from the retroreflector has its frequency shifted by Δf, where

$$\frac{\Delta f}{f} = \frac{2v}{c} \tag{10.3}$$

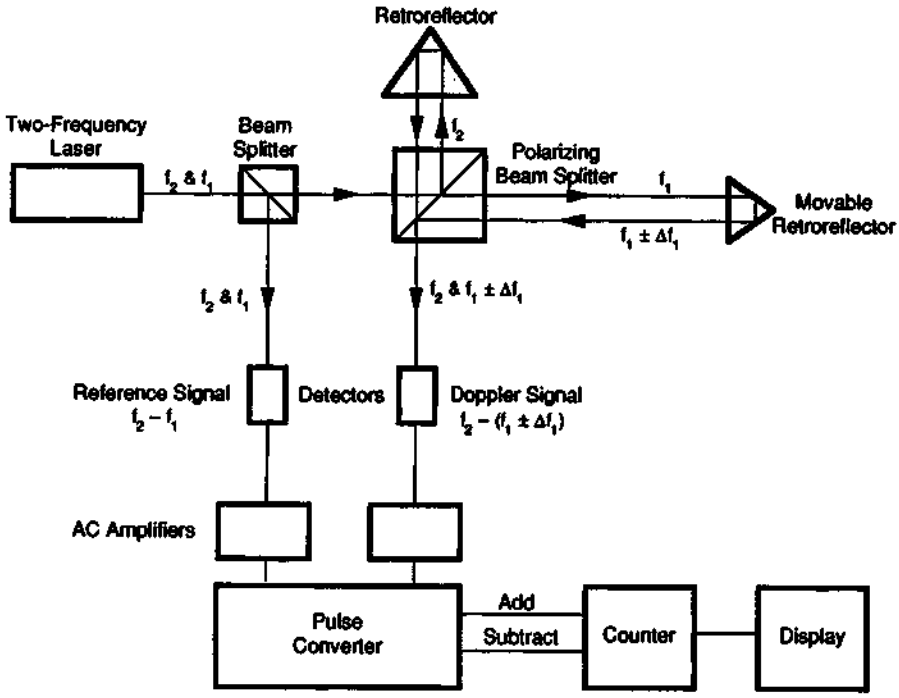


Figure 10-5 A two-frequency laser distance measuring system. f_1 and f_2 are the two laser frequencies with transverse polarizations, and Δf_1 is the Doppler shift produced by motion of the retroreflector.

where f is the laser frequency and v the velocity of the retroreflector. The velocity is integrated over time to obtain linear displacement.

A polarization-sensitive beamsplitter separates the two frequencies so that the different frequencies travel two different paths. Frequency f_2 goes to a fixed reflector, and frequency f_1 travels to the retroreflector on the part the distance of which is to be measured. The two beams are recombined at the beamsplitter and produce a beat note (amplitude modulation), as was discussed in Chapter 9. If the movable reflector moves, the returning beam will be shifted by Δf_1, so that the beat frequency will be $f_2 - (f_1 \pm \Delta f_1)$. A portion of each of the two original frequencies is used to generate a reference signal at frequency $f_2 - f_1$. The two signals go to the pulse converter, which extracts Δf_1 and produces one output pulse per quarter-wavelength of motion of the movable reflector.

Two-frequency systems can be adapted to measure pitch and yaw components of motion and straightness of travel. The operation as a straightness sensor is shown in Figure 10-6. This allows both frequencies f_1 and f_2 to reach dual retroreflecting mirrors, which are configured as indicated. Polarizing prisms separate the two frequencies because of their different polarizations. The system measures relative lateral

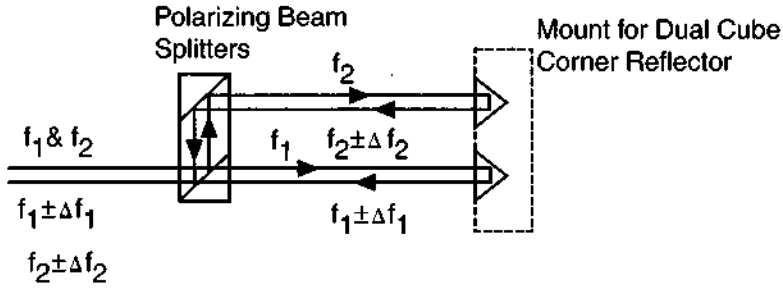


Figure 10-6 Adaptation of two-frequency system to measure straightness of travel. Here, f_1 and f_2 are the two incident laser frequencies and Δf_1 and Δf_2 are the respective Doppler shifts.

displacement between the interferometer and the axis of the reflecting mirrors. A relative lateral displacement changes the difference in optical paths between the two beams, which in turn causes a difference in the accumulated fringe counts. For a relative lateral translation x, the accumulated fringe counts will be $2x \sin(\theta/2)$, where θ is the angle between the two beams. With this adaptation, the straightness of travel of machine tools may be measured.

Operation as an angle sensor is illustrated in Figure 10-7. Parallel beams of the two frequencies f_1 and f_2 travel to the remote reflector mount, which contains two retroreflectors as shown. Angular rotation of this mirror assembly causes a Doppler shift between the return frequencies $f_1 + \Delta f_1$ and $f_2 + \Delta f_2$. The Doppler shift is not affected by axial displacement nor by small lateral displacement. The number of fringe counts accumulated is proportional to the sine of the angular rotation.

This adaptation can be used to measure the pitch and yaw components of motion, which can potentially cause errors larger than lead screw errors. It can also be used to measure flatness of surfaces, such as surface plates and machine tool beds.

Now we consider some of the sources of error in laser interferometric distance measurement. We consider first the error due to instrumental accuracy. Interferometric distance measurement systems similar to those we have described are commer-

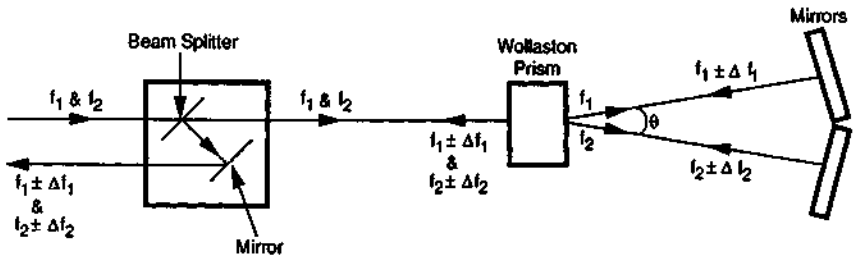


Figure 10-7 Adaptation of two-frequency system as an angle sensor. Here, f_1 and f_2 are the two incident laser frequencies and Δf_1 and Δf_2 are the respective Doppler shifts.

cially available and, in principle, can measure to an accuracy equal to a fraction of the wavelength of light. The accuracy quoted for some instruments is 1 ppm of the distance being measured, plus or minus one count. The 1 ppm is due to the transducers and the corrections for the ambient air. The error of one count represents the intrinsic resolution of the digitizing system. For many instruments, one count is equivalent to 1/4 wavelength (about 1.6×10^{-5} cm). For some multiple-path instruments, one count can be as small as 1/32 wavelength (about 2×10^{-6} cm). Thus, such interferometric systems are capable of microinch accuracy. Commercial instruments have been stated to be usable at distances up to 200 feet, although they generally are used over shorter paths. They are usually employed indoors, and sometimes in an evacuated environment.

The errors due to the atmospheric index of refraction and to material temperature effects are summarized in Table 10-2. Often, these errors are compensated for automatically and are included in the stated instrumental accuracy. The index of refraction may vary along the path of the measurement. This factor becomes more important as the measurement path increases in length.

Other sources of error arise from the particular mechanical arrangement of the measurement. These include cosine error, offset error, and preset error. These errors will be described presently.

The interferometric system measures travel in the direction of the laser beam. If the direction of the beam does not coincide exactly with the direction of travel, cosine error is introduced. If the axis of the beam is misaligned from the direction of travel by an angle α, the measured distance D_m differs from the actual distance traveled D_a by

$$D_a = D_m \cos \alpha \approx \frac{d^2}{2D_m} \tag{10.4}$$

where d is the lateral displacement at distance D_m. As an example, for a measurement at a distance of 10 feet, with an angle of 0.1° between the beam and the machine travel, the cosine error would be about 0.002 in.

Table 10-2 **Atmospheric Errors and Material Temperature Errors in an Uncompensated Interferometer**

Variable	Variation from standard conditions	Error
Atmospheric pressure	1 Torr	0.36 ppm
Atmospheric temperature	1°C	0.96 ppm
Relative humidity	1 Torr partial pressure	0.06 ppm
Workpiece temperature	1°C	$D_m K_m$[a]

[a] D_m = measured distance; K_m = coefficient of thermal expansion of workpiece material (units of cm/cm-°C).

When the travel is longer than 1 m, alignment of the laser for maximum signal over the distance of travel minimizes cosine error. Alternatively, use of a centering detector to align the beam to the direction of travel, as described in Chapter 8, can be helpful.

Offset error occurs when the beam is not in line with the tool on all axes, that is, if the object being measured (like a tool bit) is offset to the side of the laser beam. This situation is illustrated in Figure 10-8. The error arises from any rotation about an axis normal to the plane of the offset. Rotations occur because of crooked ways, thermal gradients, dirt, wear, and so forth. The magnitude d of the offset error is

$$d = A \sin \theta \tag{10.5}$$

where A is the distance from the instrument to the tool, and θ the angle at which the carriage tilts. As the figure indicates, the measured motion is D, whereas the actual motion of the tool bit is $D + d$. We note that the offset error is independent of the distance being measured. If the tilt angle is 0.1° and the displacement A is 10 in., the offset error is about 0.0017 in. This error may be reduced by eliminating the offset between the object being measured and the laser beam, if possible, and by assuring that the object being measured is parallel to the measuring system. Use of centering detectors, which give direct indications of deviation from the laser beam, can be useful.

An error called preset (or deadpath) error arises because the laser interferometer measures the distance moved from an arbitrary zero. If the interferometer is zeroed with the target at distance D_1 from the interferometer, and then the target is moved a distance D_2, the compensation for atmospheric index of refraction is applied over distance D_2. But the beam actually travels distance $D_1 + D_2$ and is affected by the changes in index of refraction over the entire path. Preset error is proportional to the preset distance and to the magnitude of the change in index of refraction. Similar considerations affect relative compensation for workpiece material temperature.

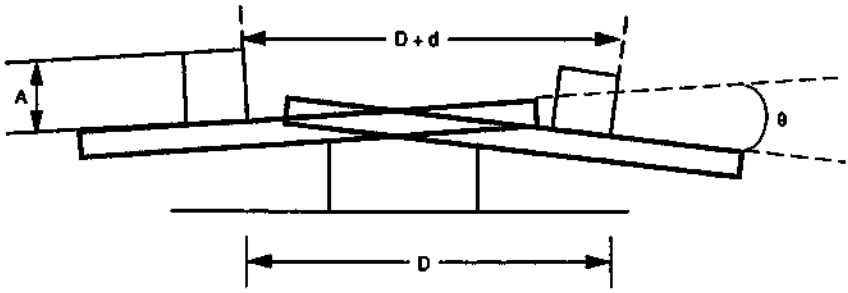


Figure 10-8 Illustration of offset error. In this figure, A is the distance from the instrument to the tool (the offset), θ is the angle at which the carriage tilts, D is the motion, and d is the offset error.

The preset error E_p can be represented by

$$E_p = D_1[0.96 \times 10^{-6}\,\Delta T - 0.36 \times 10^{-6}\,\Delta P + K_m \Delta T_m] \tag{10.6}$$

where D_1 is the preset distance (between the interferometer and the zero reference position), ΔT the change in air temperature (°C) during the measurement, ΔP the change in air pressure (Torr) during the measurement, K_m the coefficient of thermal expansion of the workpiece (cm/cm-°C), and ΔT_m the change in material temperature (°C) during the measurement. We note that preset error is present even in interferometers that compensate for atmospheric index of refraction and for workpiece temperature. This error can be compensated for by allowing the interferometer to accept preset data and correct for it. It can be minimized by positioning the interferometer optics as close as possible to the physical starting point for the motion.

As an example, at a preset distance of 30 cm, for a carbon steel workpiece with a thermal expansion coefficient of 10×10^{-6} cm/cm-°C and a change of 1°C in workpiece temperature, and with constant atmospheric temperature and pressure, the preset error would be 3×10^{-4} cm.

An estimate of the accuracy of the measurement must include all the sources of error in Table 10-3, which summarizes the foregoing discussion. It is usually helpful to compile an error budget, calculate each of the errors, and then add the individual errors to obtain a total error estimate. In differing situations, any of the individual errors in the table could be the dominant error source. In careful work, the entire error budget must be calculated to understand the accuracy of the measurement.

The maximum possible distance for laser interferometric measurements is determined by the coherence of the laser. The fringe pattern will lose contrast as one approaches a path length of the order of the coherence length S of the laser, which is

$$S < \frac{c}{\Delta \nu} \tag{10.7}$$

where $\Delta \nu$ is the linewidth of the stabilized single-frequency laser. In practice, the frequency stability of commercially available stabilized lasers is in the range 0.1–1 MHz. Hence, one is restricted to path differences less than 3×10^4–3×10^5 cm, or 300–3000 m. Practical limitations will be somewhat less.

Table 10-3 **Summary of Errors**

Error	Typical value[a]
Instrument error	1 ppm ± 1 count
Atmospheric error	Included in instrument error in compensated instruments
Material temperature error	Included in instrument error in compensated instruments
Cosine error	$d^2/2D_m$
Offset error	$A \sin \theta$
Preset error	$D_1[0.96 \times 10^{-6}\,\Delta T - 0.36 \times 10^{-6}\,\Delta P + K_m \Delta T_m]$

[a]See text for definition of symbols.

The use of laser interferometry in industrial environments will depend on the control of the environment. In practice, variations in air properties will tend to wash out the interference fringes for distances much over 100–200 feet in machine shop environments. For geodetic and seismic measurements, operation at kilometer distances has been obtained with evacuated optical paths [3]. Another practical limitation to the use of laser interferometry over long paths is the ability to determine the air parameters over the entire path and to correct them to desired reference conditions.

Laser interferometric distance measurements are useful in many industrial applications, because accuracy requirements for machines have tightened to the point that conventional linear positioning devices (like engraved line scales, magnetic pickoffs, and linear encoders) no longer provide adequate performance. To check machine positioning, one should use a measurement technique that has accuracy much better than the accuracy of the device being measured. The conventional methods just mentioned no longer offer this degree of accuracy. Specifications for machined parts may require the ability to check the accuracy of the machine tool itself to within a few parts per million. Especially when machine travel is relatively large, the measurement errors increase. Laser interferometry offers improved accuracy in many practical cases.

Laser interferometric distance measurement may be considered in the following situations:

- When the dimensional tolerances on the work approach the tolerances of the machine tool
- When the machine tool must be calibrated or checked
- When systematic deviations in the machine tool must be compensated for
- When machine tool travel covers relatively long distances

Some examples of use of laser interferometry for industrial applications include the following:

Accurate setup of work-holding fixtures for the production of aircraft engine components This application led to a considerable reduction of setup time.

Checkout of the motion of machine tools, with resultant automatic compensation for errors in tool runout An example of the use of a laser interferometer to check the motion of a machine carriage appears in Figure 10-9. This shows the position error of the motion of a jig bore machine. The cumulative error is plotted as a function of machine runout. The known error can be used to compensate and correct the runout. The measurements in this figure took a total time of 1 hour, including setup. Such measurements were previously very difficult and time consuming, especially as the runout distance grew large. The errors measured in the figure were the result of the limited capabilities of conventional systems. The use of laser distance measuring equipment can offer considerable improvement in the ability to determine and correct such errors.

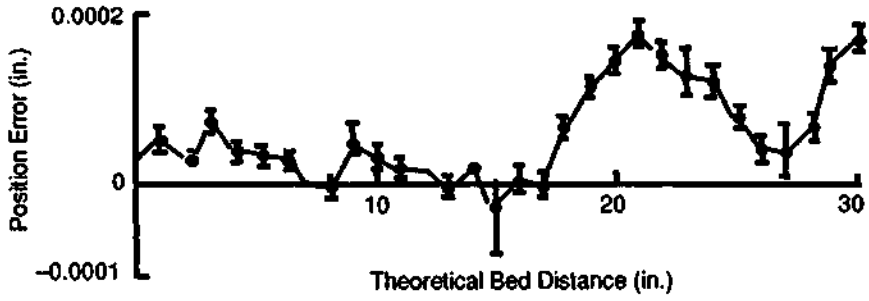


Figure 10-9 Example of the results of laser interferometric measurement of the runout of a jig bore machine, yielding cumulative position error as a function of machine travel. (Data presented by E. A. Haley at the Engineering Seminar on New Industrial Technology, The Pennsylvania State University, July 7–9, 1969.)

Speedup of positioning operations In one example, laser interferometry assisted a manual positioning system in a machining process that was otherwise numerically controlled. Aluminum plates 4 feet long required hole locations spaced at 40 in. and held to a tight tolerance, at an ambient temperature of 86°F. Automatic sensing of the part temperature coupled with the thermal expansion coefficients of the part and the machine gave 0.0005 in. accuracy for the spacings. The operator located one hole with the tool, zeroed the interferometer, retracted the tool, and manually positioned the table to the position of the next hole. The desired positional accuracy could be achieved by using the reading for hole position from the interferometer.

Rack setup on boring mills In one case involving setup of sixty-six 8 in. racks on a 44 foot horizontal boring mill, the use of laser interferometry reduced the setup time to 22 minutes, compared with a time of 11 hours when a precision peg system was used.

Building vibration measurement In one example, workers obtained a temporal record of north–south displacement of a large building in San Francisco. A laser interferometer on the ground was aimed at a retroreflector on the forty-first floor. The measurements allowed derivation of the power spectrum of the building vibration.

Measurements of strain in the earth As an example, 25-m-long interferometers near a fault in Bakersfield, CA, measured shifts of 1 part in 10^7 per day because of the motion of the fault [4]. Many other seismic and geodetic measurements of earth strain due to tides, seismic waves, and continental drift have been performed.

The foregoing examples are not an exhaustive list of possible applications for laser interferometry; they are intended to give representative situations in which laser interferometers may be used to advantage.

B. Laser Doppler Displacement

An alternative method for measuring displacement uses the Doppler shift of laser frequency. When the beam from a stabilized laser is reflected from a moving surface, the motion of the target induces a frequency shift of the incident beam. This frequency shift may be measured and converted to a measurement of surface displacement.

Compared with interferometric measurements, the laser Doppler displacement method offers similar accuracy and resolution, but potentially lowered cost and simplified setup. It has similar restrictions with respect to indoor use and maximum measurement distance as the interferometric methods. Some characteristics were presented in Table 10-1.

Commercial models of devices using laser Doppler displacement are available. They measure the light that is reflected by a retroreflector attached to the moving part the distance of which is to be measured, just as interferometric devices do. But instead of measuring fringe counts, they measure a phase shift that corresponds to the displacement of the retroreflector. The measurement accuracy for these devices is claimed to be in the microinch range. Quoted accuracy ranges from 10^{-7} to 10^{-8} of the distance being measured. The laser Doppler displacement meters also offer environmental compensation, including corrections for atmospheric temperature and pressure and for thermal expansion.

As with interferometric devices, there are variations that allow measurement of straightness, angular orientation, flatness, and squareness. Applications have included calibration of machine tool runout, calibration of flatness of surface plates, measurement of the pitch and yaw angles of *XY*-tables, and measurement of the straightness of machine motion.

C. Beam Modulation Telemetry

For measurements of large distances in the field, interferometric and Doppler displacement methods with microinch sensitivity are no longer applicable. Fluctuations of the density of the atmosphere over paths exceeding a few hundred feet make interferometry impractical.

A common method of distance measurement outdoors over long distances involves amplitude modulation of the laser beam and projection of the modulated beam toward a target. The lasers most often used have been helium–neon lasers or semiconductor diode lasers. The target may be a retroreflector positioned at the spot the distance of which is desired. Alternatively, the target may be a natural feature the distance of which is to be measured. Measurements have been made using many different types of surfaces (buildings, water, grass). The use of a natural feature as the reflector allows a single-ended measurement and does not require positioning of a retroreflector, but the received signal will be smaller than if a retroreflector were used.

The reflected return from the target is collected by a telescope and sent to a detector. The phase of the amplitude modulation of the returning light is compared with the phase of the amplitude modulation of the emitted light, A difference in phase occurs because of the finite time for the light to travel to the target and return to the telescope. The phase shift ϕ is related to the total path length L by the equation

$$\phi = 2\pi \frac{2n_g L}{\lambda_v} \tag{10.8}$$

where λ_v is the vacuum wavelength and n_g is the group index of refraction of the air. The group index will be discussed presently.

To determine distance, one measures the phase shift of the collected light relative to the outgoing light and then uses Equation (10.8). Figure 10-10 shows a schematic diagram of a system for such a measurement. The light is amplitude modulated at a known frequency, collimated, and transmitted to a distant reflector. Reflected light is collected by the telescope and focused on a detector. A phase detector compares the relative phase of the amplitude of the return beam with that originally present.

The phase difference can be represented as

$$\phi = (N + f)2\pi \tag{10.9}$$

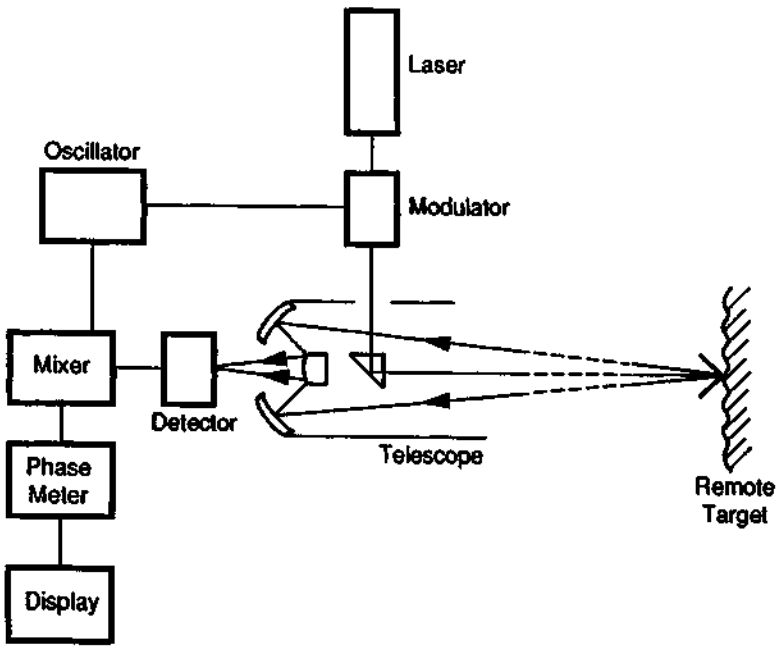


Figure 10-10 Schematic diagram of distance measurement using beam modulation telemetry.

where N is an integer and f is a fractional part less than unity. The original measurement gives f, but no information about N. Thus, there is an ambiguity in the phase measurement by a factor of an integral multiple of 2π.

To determine N, one may make several measurements at different values of the modulating frequency. One begins at a low value of frequency, such that $\phi < 2\pi$ (i.e., $N = 0$), and obtains a course measurement. Then the frequency is increased and f determined more accurately, with N known from the earlier course measurement. In several steps, one may obtain ϕ with high accuracy.

Another variation allows changing the modulation frequency so as to obtain a null output when the original signal and reflected signal are 180° out of phase. If the frequency is varied, nulls will be obtained for variations in the optical path corresponding to one-half wavelength of the modulation frequency. Measurements of several nulls will provide enough information to determine the distance without ambiguity. Because the frequency at which a null occurs may be measured accurately, this method offers higher precision than direct measurement of a nonzero phase shift.

The limitation to the accuracy of the distance measurement is often knowledge of the proper value of the index of refraction of the air through which the measurement is made. The index of refraction used to correct for atmospheric effects is the so-called group index, which is defined by

$$n_g = n + \sigma \frac{dn}{d\sigma} \tag{10.10}$$

with σ the reciprocal of the vacuum wavelength and n the phase index of refraction, which is the usual index we have been using. The difference arises because in dispersive media (that is, media in which the index is a function of wavelength), the group velocity (the velocity with which the energy of the wave propagates) differs from the phase velocity (at which the phase of the wave propagates). In interferometric measurements, as previously described, the phase index is the appropriate index to use; but in beam modulation telemetry, the group velocity and group index are the appropriate parameters. The value of the group index at the helium–neon laser wavelength (0.63299138 μm in vacuum) is

$$n_g = 1.0002845073 \tag{10.11}$$

for dry air at 15°C, 760 Torr, and 0.03 percent carbon dioxide content. Corrections for the group index as a function of varying atmospheric parameters are available in the literature [5].

Determination of the proper corrections to be applied to the index of refraction in the field is difficult. One must actually use a value of n_g averaged over the entire path traversed by the light. If the terrain is level and meteorological conditions are stable, the average may be relatively easy to obtain. One makes measurements of temperature and pressure at a number of points along the path and averages them. But this can involve extra labor and still not yield complete accuracy, especially if the air is unstable or turbulent. If the terrain is difficult, it may not be practical to

measure at many points along the path. An example would be mountainous terrain. A method for averaging the index of refraction using two different optical wavelengths has been suggested [6]. In tests using this method over a 1 mile path at the Stanford linear accelerator, a reproducibility of 1 part in 10^7 was observed.

Beam modulation telemetry units have been developed for which the claimed accuracy is ±1 mm at distances up to approximately 1000 m and 1 part in 10^6 for greater distances. Such devices are employed in a variety of applications, including land profiling and geodetic survey work. An example is the accurate measurement of the relative motion of the blocks making up the opposite sides of the San Andreas fault in California. Beam modulation telemetry has also been used for measurement of the motion of large structures, such as bridges and dams. The system can be employed as an altimeter and used for purposes such as slope surveys and gravimeter calibration.

When the system is used for applications like altimetry or measurement of the distance to a topographic feature, it is usually used as a single-ended system, operated by one person, and using the reflection off the target for the return signal. When it is used for applications like surveying or geodetic work, it often uses a second operator who positions a retroreflector.

Competing methods for measurements that are performed by beam modulation telemetry include microwave radar and the use of geodetic measuring tapes. Microwave radar used in this type of application suffers from problems arising from ground clutter. The use of geodetic tapes requires a large amount of time for accurate measurement.

A dramatic example of the use of a laser beam modulation distance measuring system occurred in mapping of the Grand Canyon [7]. The mapping provided accurate profiles for a large-scale map of a 310 square mile area in Grand Canyon National Park. The survey was completed in three days by two people and provided improved maps of the park area. It was estimated that conventional taping methods would have taken one year and required 100 people. Other, more routine, surveying tasks are also speeded by the use of such laser units.

One specific example of the unusual capabilities of laser beam modulation telemetry involves measurement of ocean wave profiles. Figure 10-11 shows an

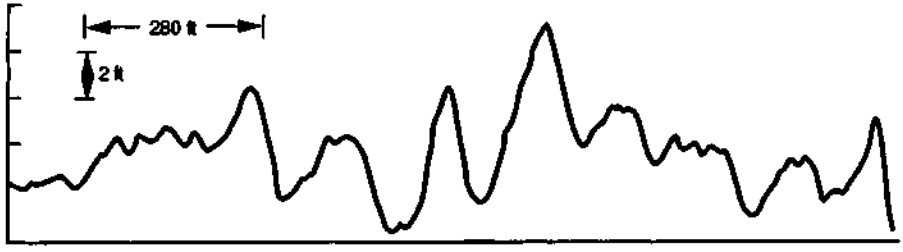


Figure 10-11 Ocean wave profile obtained using beam modulation telemetry with a laser flown in an airplane at an altitude of 500 feet.

example of the wave profile obtained 20 miles off the coast of New Jersey using a laser device flown in an airplane at an altitude of 500 feet. The horizontal resolution was 4 in., the diameter of the laser beam at the water surface. The figure shows graphically the capability of laser distance measuring systems for making difficult types of measurements.

D. Pulsed Laser Range Finders

Another approach to laser-based distance measurement involves measurement of the round-trip transit time for a very short pulse. Such devices have commonly used Q-switched lasers with high peak power. The most common uses have been in satellite tracking and in military range finding. Such applications are outside the scope of this book, but the technique does have a variety of other engineering and commercial uses. These applications include generation of accurate terrain maps, calibration of aircraft radar, and measurements of the ranges of aircraft and ground vehicles.

Pulsed laser range finders operate on the principle of emitting a very short pulse of laser light and measuring the transit time for the pulse to reach a visible target and for the reflected pulse to return to a receiver located near the laser. This approach is sometimes called an optical radar or LIDAR, which stands for LIght Detection And Ranging. The system consists of a laser, a telescope to collect the reflected light, a photodetector, and a precision clock to measure the difference in time between the outgoing pulse and the return pulse. The narrowly collimated laser beam makes it possible to measure the range of specific targets and to receive a return free from ground clutter or interference from other objects near the target.

Because the target must be visible, the range of this type of measurement is limited by atmospheric conditions. Again, the main limitation to the accuracy of the measurement is lack of exact knowledge about the index of refraction of the air through which the laser pulse passes.

These measurements are commonly called cooperative if the target is equipped with a reflective surface or a retroreflector to enhance the return signal. They are called noncooperative if the target surface has not been prepared so as to enhance the return signal. Ranging to topographic features or military range finding is usually of the noncooperative type.

For military uses, pulsed distance measurement is preferred to other laser-based measurements because the range is obtained very quickly. The Q-switched Nd:YAG laser has been used most commonly for this application. Laser range finders have been used by aircraft, tanks, and artillery and are routinely used to obtain ranges by all the world's major military organizations.

An example of the use of lasers for satellite tracking involves the NASA Laser Geodetic Satellite (Lageos), launched in 1976. Lageos is a passive satellite, with a surface studded with 426 fused silica retroreflectors. It is illuminated by Nd:YAG laser pulses with duration around 200 psec. The ground-based laser stations transmit pulses to the satellite and receive the return signals from the retroreflectors. The resulting range data are used to compute precise coordinates for the satellite and for

the ground stations. Space geodesy has been used to determine the coordinates of the stations and their relative separations with centimeter accuracy. Among other things, tectonic motion and deformation have been determined to high precision [8].

Another long distance application of pulsed laser range finding has been measurement of the distance to the moon. The round-trip transit time to retroreflectors left by Apollo astronauts has been measured. Panels containing a number of retroreflectors were deployed on the Apollo 11, 14, and 15 missions, and also by the unmanned Soviet Luna 17 and 21 missions. At the McDonald Observatory of the University of Texas, a Q-switched ruby laser pulse with a duration around 4 nsec was transmitted through a 2.7 m telescope [9]. The lunar distance was obtained from the transit time of the return pulse to an estimated accuracy of 15 cm [10]. Data from these measurements was used to generate a new lunar ephemeris and to check scientific theories about the lunar orbit and lunar mass distribution.

Commercial models of pulsed laser distance measuring equipment are available from a number of sources. Models include devices based on pulsed semiconductor diode lasers, on diode-pumped solid state lasers, and on flashlamp-pumped solid state lasers. The maximum range that can be measured for a low-reflectance target varies from a few hundred meters for the diode laser models, to a few kilometers for the diode-pumped solid state models, to a few tens of kilometers for flashlamp-pumped solid state models. The maximum ranges increase substantially if a retroreflector is used as the target. At least one model of a handheld portable device based on a semiconductor diode laser is available.

Quoted measurement accuracies are a few millimeters to 1 cm, if averaging is performed over a number of pulses. Figure 10-12 shows an example of a commercial flashlamp-pumped Nd:YAG ranging system.

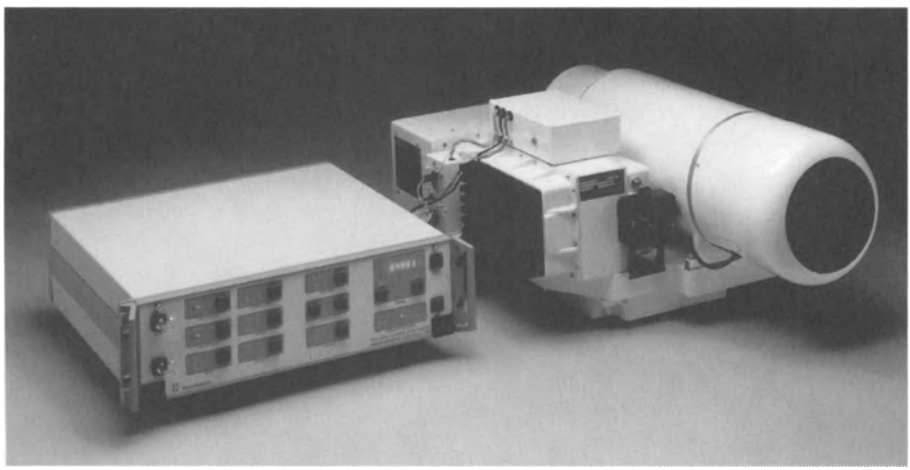


Figure 10-12 Commercial laser ranging system based on a flashlamp-pumped Nd:YAG laser. (Photograph courtesy of Azimuth Corp.)

E. A Laser Interferometer Application in Mask Production: A Specific Example of Distance Measurement and Dimensional Control

In this section we present a specific example of the use of a laser interferometer for distance measurement and dimensional control in a photolithographic mask making process. The laser is a two-frequency helium–neon laser with an output around 1 mW. The system employs the type of measurement described in Figure 10-5.

The system has two applications, one as a pattern generator and the second as a step-and-repeat camera for production of copies of a pattern. In the pattern generation mode, the workpiece is moved to a previously determined position. Movable apertures are positioned to form a rectangular area, which is to be exposed. The workpiece has photosensitive material, which is exposed by pulsing an incoherent light source. The laser interferometer measures and controls five degrees of freedom: two dimensional coordinates of the center of the rectangle, the length and width of the rectangle, and the angular orientation of the rectangle. Under the control of a computer, the rectangles are automatically exposed sequentially in order to form the desired pattern, which is then processed photolithographically.

In the step-and-repeat mode, a previously generated pattern is used as a reticle. The system moves one step and repeats exposures with a ten-to-one reduction in size. In this mode, the laser interferometer controls the two positional coordinates only.

The interferometer has also been used to measure the straightness of the runout of the mechanical system, the pitch and yaw of the mechanical motion, and the flatness of the surface in which the positional measurements are made. All these parameters must be controlled to high accuracy in order to achieve acceptable results. The laser system attains an accuracy of ± 0.25 μm in the positional coordinates and 0.1 arc sec in the rotation coordinate.

The system is employed for production in an enclosed room in which the temperature and relative humidity are controlled. Changes in air pressure are monitored and compensated for. The system is employed because it offers greater accuracy and tighter control than competing techniques, such as mechanical indexing systems or stepping motors. The system is routinely used for highly accurate generation of mask patterns.

References

[1] J. C. Owens, *Appl. Opt.* **6**, 51 (1967).
[2] J. N. Dukes and G. B. Gordon, *Hewlett-Packard J.*, p. 2 (August 1970).
[3] V. Vali and R. C. Bostrom, *Rev. Sci. Instrum.* **39**, 1304 (1968).
[4] P. L. Bender, *Proc. IEEE* **55**, 1039 (1967).
[5] J. C. Owens, *Appl. Opt.* **6**, 51 (1967).
[6] J. C. Owens, Laser Applications in Metrology and Geodesy, in *Laser Applications I* (M. Ross, ed.). Academic Press, New York, 1971.

[7] *Opt. Spectra*, p. 12 (May 1973).
[8] D. E. Smith, *et al.*, *J. Geophys. Res.* **95**, 22013 (1990).
[9] P. L. Bender, *et al.*, *Science* **182**, 229 (1973).
[10] E. C. Silverberg, *Appl. Opt.* **13**, 565 (1974).

Selected Additional References

A. Interferometric Distance Measurement

R. R. Baldwin, G. B. Gordon, and A. F. Rudé, Remote Laser Interferometry, *Hewlett-Packard J.* (December 1971).
C. W. Gillard and N. E. Buholz, Progress in Absolute Distance Interferometry, *Opt. Eng.* **22**, 348 (1983).
C. W. Gillard, N. E. Buholz, and D. W. Ridder, Absolute Distance Interferometry, *Opt. Eng.* **20**, 129 (1981)
J. T. Luxon and D. E. Parker, eds., *Industrial Lasers and Their Applications*, Prentice-Hall, 2nd edition, Englewood Cliffs, NJ, 1992, Chapter 10.

B. Laser Doppler Displacement

C. P. Wang, Laser Doppler Displacement Measurements, *Lasers and Optronics*, p. 69 (September 1987).

C. Beam Modulation Telemetry

K. B. Earnshaw and J. C. Owens, A Dual Wavelength Optical Distance Measuring Instrument which Corrects for Air Density, *IEEE J. Quantum Electron.* **QE-3**, 544 (1967).
B. E. Edwards, Design Aspects of an Infrared Laser Radar, *Lasers and Applications*, p. 47 (October 1982).
A. Greve and W. Harth, Laser-Diode Distance Meter in a KERN DKM 3A Theodolite, *Appl. Opt.* **23**, 2982 (1984).

D. Pulsed Laser Range Finders

C. G. Bachman, *Laser Radar Systems and Techniques*, Artech House, Dedham, MA, 1979.
G. Beheim and K. Fritsch, Range Finding Using Frequency Modulated Laser Diode, *Appl. Opt.* **25**, 1439 (1986).
H. N. Burns, C. G. Christodoulou, and G. D. Boreman, System Design of a Pulsed Laser Rangefinder, *Opt. Eng.* **30**, 323 (1991).
D. W. Coffey and V. J. Norris, YAG:Nd^{3+} Laser Target Designators and Range Finders, *Appl. Opt.* **11**, 1013 (1972).
J. J. Degnan, Satellite Laser Ranging: Current Status and Future Prospects, *IEEE Trans. Geosci. Remote Sensing* **GE-23**, 398 (1985).
P. Hermet, Design of a Rangefinder for Military Purposes, *Appl. Opt.* **11**, 273 (1972).
A. V. Jelalian, *Laser Radar Systems*, Artech House, Boston and London, 1992.

E. A Laser Interferometer Application in Mask Production: A Specific Example of Distance Measurement and Dimensional Control

E. A. Hilton and D. M. Cross, Laser Brightens the Picture for IC Mask-Making Camera, *Electronics*, p. 119 (August 7, 1967).

Chapter 11 | Laser Instrumentation and Measurement

The unusual properties of laser light can be used in a variety of ways for industrial applications. We have already discussed the application of lasers for distance measurement in the previous chapter. In this chapter, we will discuss other applications involving measurement.

A wide variety of inspection and measurement procedures may be accomplished easily using lasers. Many such devices have been developed as commercial instrumentation. For other applications, the instrumentation is developed by the individual user to meet specific needs.

The variety of different measurement applications is so great that it is impossible to give a complete summary. The selection in this chapter will show the great adaptability of laser-based instrumentation for conducting unusual types of measurements, and its capabilities for making measurements that might be difficult with conventional techniques. In particular, we discuss methods for measuring velocity (both of fluid flow and of solid surfaces), for measuring angular rotation rate, for measuring the diameter of wires and fibers, for determining surface profile and position, for measuring dimensions of finished products, and for particle sizing. The diverse measurement capabilities summarized here can serve as examples for applying similar techniques to one's own measurement problems.

A. Velocity Measurement

1. MEASUREMENT OF FLUID FLOW

Measurement of the velocity of fluid flow can be performed by scattering a laser beam from a liquid or gas. The laser beam interacts with small particles carried along by the flowing fluid. The particles scatter some of the light in all directions. Many fluids contain enough particles so that the scattered light can easily be detected.

The frequency of the scattered light is slightly shifted by the Doppler effect (see Chapter 9). The magnitude of the frequency shift is proportional to the velocity of the fluid. Measurement of the frequency shift directly gives the flow velocity, or at least one component of it. Laser Doppler velocimetry has been employed in many scientific and engineering applications to measure flow velocity. It is a remote, noncontact method that is especially applicable in many situations where other measurement techniques are difficult. Commercial laser-based instrumentation for measurement of flow velocity is available from a number of sources.

One of the greatest advantages of laser Doppler velocimetry is that there is no physical contact with the fluid. It is not necessary to introduce any probes or other equipment into the flow; hence, the flow is not disturbed. Hot or corrosive fluids can be measured without problems. Measurements can be conducted at some distance from the flowing fluid, if necessary. High-resolution measurements may be performed on complex flow fields.

The narrow spectral width of the laser light makes accurate measurements possible. The Doppler shift of the scattered light can be greater than the laser linewidth, even for relatively slow flow velocity. This fact gives the laser velocimeter a potentially wide dynamic range. It is possible to measure flow velocity ranging from centimeters per second to hundreds of meters per second.

The speed of the response can be high. This is in contrast to most conventional flow measuring devices, which have more sluggish response. This makes it possible to perform measurements related to transient conditions and to investigate turbulent flow characteristics.

The main disadvantage of laser Doppler velocimetry is the necessity of having scattering particles entrained in the flow. Particles are needed to provide sufficient scattered light to yield a detectable signal. In many practical circumstances, fluids of interest have sufficient scattering centers to provide a measurable output. In cases where the fluid is very clean, with no scattering centers, it may be possible to seed the flow with scattering particles. Particles seeded into the flow must be very small so as to follow the flow faithfully.

Laser Doppler velocimetry is based on the Doppler shift. The frequency of the light scattered by an object moving relative to a radiating source is changed by an amount that depends on the velocity and the scattering geometry. The measurement technique basically consists of focusing laser light at a point within the flowing fluid. Light scattered from the fluid or from particles entrained within the fluid flow is collected and focused on an optical detector. Signal processing of the detector output yields the magnitude of the Doppler frequency shift and, hence, the velocity of the flow.

Various optical arrangements have been used. One common approach is illustrated in Figure 11-1a. This arrangement is called a dual beam mode. Light from a continuous laser is split into two equal parts by a beam splitter. The lens focuses the beams to the same position in the fluid. The operation of the dual beam system can be understood in terms of fringes. Where the two beams cross in the fluid, they interfere to form fringes consisting of alternating regions of high and low light intensity. The

fringe pattern is illustrated in Figure 11-1b. When a particle traverses the fringe pattern, it will scatter light when it is passing through regions of high intensity. The scattering will be reduced when the particle is passing through regions of low intensity.

Light scattered by a particle in the fluid and arriving at the detector will produce a varying output, the frequency of which is proportional to the rate at which the particle traverses the interference fringes. The distance s between fringes is

$$s = \frac{\lambda}{2 \sin \frac{\theta}{2}} \tag{11.1}$$

where θ is the angle between the two converging beams and λ is the wavelength. If a particle passes through the fringes with a component of velocity $v_\perp$ in the direction perpendicular to the fringes (and also approximately perpendicular to the light beams), the output signal from the detector will be modulated at a frequency f given by

$$f = 2v_\perp \frac{\sin(\theta/2)}{\lambda} \tag{11.2}$$

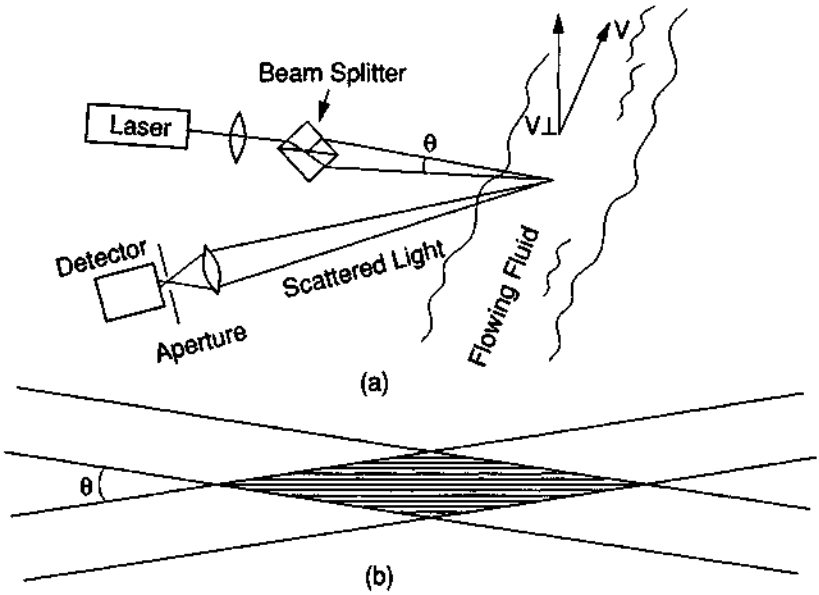


Figure 11-1 (a) Laser fluid velocimetry. (b) The pattern of interference fringes formed by the two beams that intersect at a small angle θ. The fluid flows with velocity v, with a component $v_\perp$ perpendicular to the laser beam.

The wavelength λ is the wavelength in the fluid, which differs from the vacuum wavelength λ_0 by a factor equal to the index of refraction n of the fluid. Thus, Equation (11.2) may be rewritten

$$f = 2nv_{\perp} \frac{\sin(\theta/2)}{\lambda_0} \tag{11.3}$$

The collecting optics are arranged to view the point of intersection of the two beams in the fluid. The detector receives light scattered from this region. This allows the system to be single-ended, that is, the detector and the laser are near the same position. This is the most convenient arrangement when remote measurements are to be made at a distance. When the fluid sample is nearby and accessible, the collector may be located on the far side of the fluid from the laser, and light scattered in the forward direction will be collected. The use of forward scattered light increases the signal and reduces requirements on laser power.

The required laser power depends on the geometry and the distance at which the measurement is to be made. For measurements made at short distances (perhaps 10–20 cm), helium–neon lasers with outputs in the milliwatt range are adequate. When measurements are to be made at larger distances (for example, remote anemometry), argon lasers with outputs of 1 W or greater are commonly used.

The arrangement shown in Figure 11-1 detects one component of the velocity vector. To characterize the velocity vector completely in the general case, one would have to use three noncoplanar laser sources.

The dual beam arrangement shown in Figure 11-1 forms the basis for operation of many laser Doppler velocimeters and laser Doppler anemometers. Many variations on this basic scheme are possible, including the simultaneous use of two laser colors or two laser polarizations to determine two components of the velocity in the flow field.

Early instrumentation, which provided many experimental measurements and which is described in many technical articles, used a so-called reference beam mode. In this mode, the scattered light is collected from particles going through a single beam, and it is mixed in the detector with unscattered light directly from the laser. The scattered light is frequency shifted by the frequency given in Equation (11.3). The frequency shift is detected as a beat note in the output of the detector. This point of view yields exactly the same result as consideration of a particle traversing a fringe pattern in the dual beam mode.

The reference beam mode requires that the aperture through which the scattered light is collected be small. Thus, the signal is smaller than that available in the dual beam mode, and laser power requirements are higher. The dual beam mode of operation appears to be preferable when the number of scattering particles is small. If the particle density is high and many particles are traversing the measurement volume at the same time, the reference beam mode may be preferable.

Many different measurements may be made. One can measure the stream velocity in a given direction, that is, one component of the velocity at a fixed position.

One may change the position of the measurement region and map out profiles of steady flows. With added complexity, one may measure two or three components of velocity at a fixed position, or one may determine the complete flow field of a steady flow. If the velocity is varying with time, because of turbulence, the frequency will be a rapidly varying function centered on a mean value. Measurement of the frequency spectrum can provide information about the spectrum of the turbulence. One may use a spectrum analyzer with frequency tracking capability to provide instantaneous information about the turbulence at a specified point in the fluid.

Laser Doppler velocimeters have been employed for many types of flow studies. Applications have included study of laminar and turbulent flows, investigation of boundary and shock phenomena, determination of three-dimensional velocity data in the vortices near the tips of aircraft wings, measurement of flow between the blades of turbines, gas velocity measurements within compressor rotor passages, in vivo measurement of blood flow, measurements of flow velocity of chemical reactants for process control in production applications, and remote measurement of wind velocity.

To demonstrate the capabilities of laser velocimetry, we present two specific examples. Figure 11-2 shows the velocity profile of laminar flow for water through a circular pipe, in which a constant flow velocity of 5.35 cm/sec was maintained near the center of the tube. One expects the velocity distribution of a viscous fluid undergoing laminar flow in a pipe to be parabolic, with the velocity being maximal at the center and zero at the walls. The measurements shown in this figure demonstrate the expected behavior. They also show the versatility of laser velocimetry in probing a flow profile, without introducing probes that could affect the flow.

Figure 11-3 shows a longitudinal velocity profile of a shock wave resulting from a supersonic jet of nitrogen gas ejected under pressure into ambient air through a supersonic nozzle. A wedge set in the flow stream produced the shock wave. The longitudinal velocity scan was made 0.15 in. above the centerline of the flow. The results show how one can make a direct measurement of flow velocity throughout a shock wave without introducing any perturbations into the flow. The figure demonstrates high spatial resolution for the measurement. It would have been difficult to obtain this measurement using conventional techniques. These two examples illustrate the wide range of fluid velocity measurements possible using laser techniques.

Laser velocimeters have been employed to detect wind shear near airports. They may offer superior performance to Doppler radar for detection of wind shear under dry conditions. This application has not yet been adopted for routine monitoring of wind shear.

Laser devices have been extensively studied for air data sensing on board commercial and military aircraft. The usual method of detecting air speed data has been the use of pitot tube air pressure sensors. Remote laser sensing could provide measurements of air data outside the boundary layer in the air surrounding the aircraft. Also, it should provide superior performance under conditions where the angle of attack of an aircraft is high. This application appears promising for future implementation, but it has not yet been adopted for routine use.

Commercial laser instrumentation is available in a wide variety of forms, including devices based on helium–neon and on argon lasers, devices using forward scat-

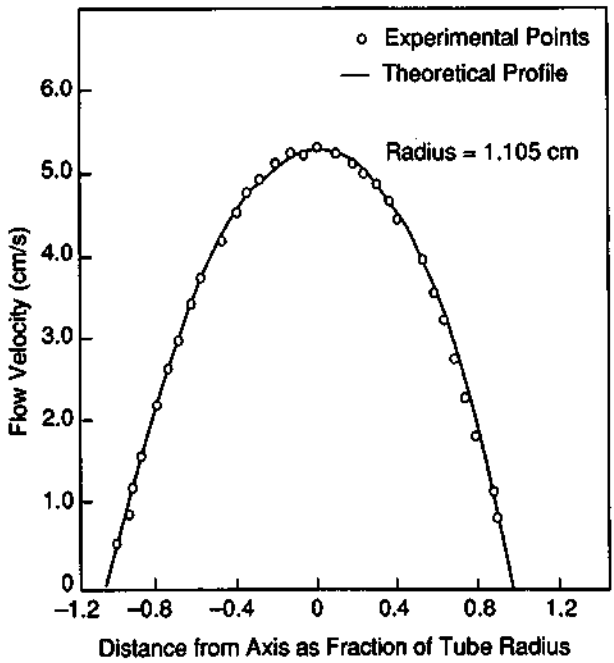


Figure 11-2 Velocity profile of laminar water flow through a circular pipe. (From J. W. Foreman *et al.*, *IEEE J. Quantum Electron.* **QE-2**, 260 (1966).)

tering and backscattering, and devices capable of measuring one, two, or three components of fluid velocity. Quoted measurement accuracy of 0.5 percent at 40 MHz is available. Such devices provide capability for making many routine fluid flow measurements and also for making unusual types of measurements that would be difficult or impossible using conventional techniques.

An alternative method for determining complex flow fields in a single measurement involves particle image velocimetry (PIV). Velocity is determined simultaneously at many points in a two-dimensional plane. The beam from a pulsed laser is spread into a sheet by lenses and illuminates a plane in the flow field. An image is recorded of the particles entrapped in the flow. After a delay, which may range from microseconds to seconds, the laser is pulsed a second time, and the image of the particles is recorded again. The images are digitized and processed to determine the displacements of the particles in the time between the pulses. The flow velocity is derived from the displacements of the particles.

Because PIV determines the entire structure of the flow field in one measurement, it provides detailed information about spatial structures in the flow and is useful for visualization of flow patterns. It has been used for studying unsteady

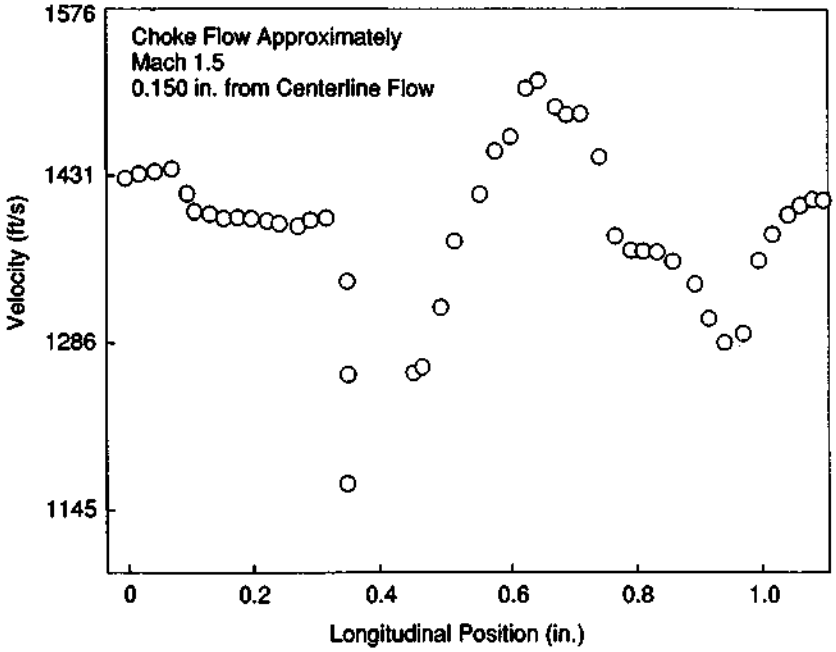


Figure 11-3 Longitudinal velocity profile of a shock wave for a supersonic jet of nitrogen ejected into ambient air through a supersonic nozzle. (From A. E. Lennert *et al.*, at the 2nd National Laser Industry Association Meeting, Los Angeles, October 20–22, 1969.)

aerodynamic flows and turbulent water flows, for aircraft design, and for determining flow structures in combustion processes.

Commercial models of PIV equipment are available, using both 532 nm Nd:YAG lasers and pulsed argon lasers. The commercial models use various techniques to record the images and process the data. They can measure flow velocity in the range from 0.00025 to 1250 m/sec. Sophisticated displays of the flow field are available.

2. *VELOCITY OF SOLID SURFACES*

The velocity of solid surfaces may be measured by methods similar to those discussed in the previous section. The measurements use the Doppler shift of laser light reflected from the surfaces. The principle is the same as that used for measuring fluid velocity, except that the frequency shift is generated by reflection of light from a moving solid surface instead of by scattering in a flowing fluid. Such measurements have been used to measure diverse types of motion, such as vibrations in large structures, velocity of moving semimolten fibers in the polymer industry, speed of

rolling mills in production of metals, web speed in paper mills, and the velocity of magnetic tape.

One example that illustrates the versatility of laser-based solid surface measurement involves determination of the motion of the surface of metallic plates that have been shocked by the impact of a high-power pulse from a CO_2 TEA laser. The configuration is shown in Figure 11-4. The multimegawatt CO_2 laser pulse is incident on the plate from the left. It ablates material from the surface and generates a shock wave that travels through the plate. When it reaches the right surface, that surface moves in response to the shock. The right portion of the apparatus constitutes a Michelson interferometer. The detector views the motion of the fringes as the surface responds to the shock. The velocity of the shock wave is obtained from the time between the CO_2 laser pulse and the beginning of the motion of the fringes. The amplitude of the motion can be derived from counting the number of fringes that pass over the detector. This information, coupled with the equation of state for the plate material, yields the pressure in the shock wave. Finally, the magnitude of the shock as a function of CO_2 laser energy can be derived. This example shows how lasers can be used to perform measurements that would be difficult with other techniques.

Another type of laser velocimeter for solid surfaces has been developed that operates on an entirely different principle [1]. This velocimeter is well suited for measuring velocity of solid materials with relatively rough surfaces. The velocimeter just described was better suited to a solid with a relatively shiny surface, so that a reasonably well-collimated beam of light would be reflected from the surface.

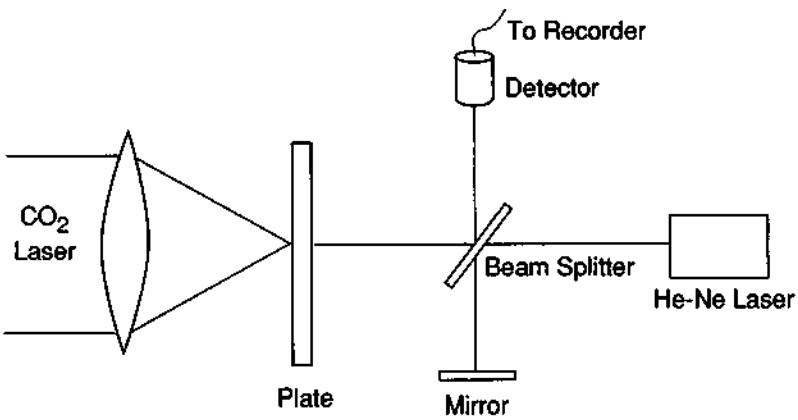


Figure 11-4 A laser velocimeter for determination of the motion of the surface of a metal plate resulting from CO_2-laser-generated shock wave propagation through the plate.

The second velocimeter uses the so-called "speckle pattern" that appears in laser light reflected from a diffusely reflecting surface. An observer viewing such a reflection sees an illuminated area covered by tiny specks of light. This appearance is caused by interference of the coherent laser light. Light reflected from small irregularities present on the diffuse surface can interfere to produce regions with a high local light intensity. The observer sees these regions as bright speckles. Such constructive interference effects are also present when a surface is illuminated by conventional incoherent light sources, but the pattern of the speckles changes randomly in a time short compared with the resolving time of the eye. Thus, a surface illuminated by a conventional light source appears uniformly lighted, whereas a diffuse surface illuminated by coherent laser light appears to have a granular speckled texture.

When there is relative motion between the surface and the observer, the speckles appear to move. The entire pattern of bright spots moves as a whole. The apparent velocity of the speckles is twice the velocity of the relative motion.

If the eye of the observer is replaced by a photodetector with a small aperture, movement of one speck of light over the aperture will result in a pulse from the detector. This yields a method of measurement for the surface velocity. An apparatus for quantitative measurement is shown in Figure 11-5. The speckle pattern is shown schematically as an irregular pattern of reflected light. A grating is positioned in front of the detector. As one bright spot in the speckle pattern moves over the grating, it will produce a series of pulses. The frequency f of the series of pulses will be

$$f = \frac{2V}{d} \tag{11.4}$$

where V is the velocity of the surface and d is the grating spacing. There will be a large number of speckles crossing the grating. Each will give a series of pulses. The pulse trains from different specks will have random phases, but they will all have the same frequency. Thus, it is only necessary to measure the center frequency of the output of the detector to determine the surface velocity.

Instruments based on this principle can provide a noncontact sensing element suitable for use on surfaces with any color or texture. The measurement method has been used successfully on a wide variety of materials, including aluminum, black and shiny steel, copper wire, newsprint, clear and opaque plastic, leaves, sand, and dirt. The accuracy of the measurement is stated to be 0.1 percent, over the range of velocities from 50 to 1000 ft/min. It can be used to measure the velocity of vehicles with greater accuracy than a fifth wheel. It is especially suited for use with automated processes involving continuous measurement of speed or length of products. It has been used, for example, for measuring the speed of strip metal in steel and aluminum mills.

These two illustrative examples do not exhaust the possible applications of lasers for measurement of solid surface velocity, but they do demonstrate the broad capability of laser sensing for velocity measurement.

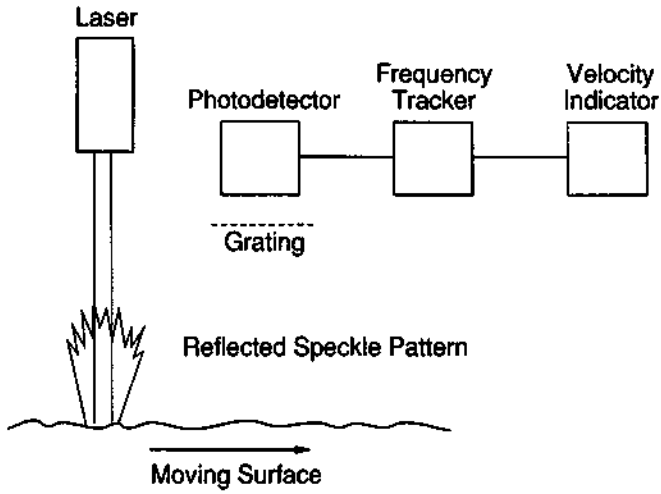


Figure 11-5 Apparatus for measuring solid surface velocity using the speckle pattern of reflected laser light.

B. Angular Rotation Rate

Lasers may be used for sensing of angular rotation rate, an application that makes possible a laser gyroscope. The so-called ring laser gyro uses a ring configuration, usually a triangle, in which two counterrotating laser beams travel in opposite directions around the ring. The wavelength of operation of the laser gyro adjusts itself so that the distance around the ring is an integral number of wavelengths. A change in the length of the ring causes a change in the wavelength. If the ring rotates about a direction perpendicular to the plane of the ring, there will be a change in the effective length of the paths seen by the two counterrotating beams. The beam traveling in the same direction as the sense of rotation will have to travel slightly farther on each traversal of the ring. The system is shown schematically in Figure 11-6, which shows a triangular arrangement. The active laser medium is in two legs of the triangle. Mirrors at the end of each leg form a closed resonator that defines the paths of the beams.

The velocity of light is very large, but not infinite. During the time that the light wave makes one transit around the ring, the ring will move slightly. Thus, one of the two beams must travel slightly farther to traverse the ring. This changes the effective distances that the beams travel around the ring in the two directions. The frequencies of the two beams are determined by the optical path lengths. The number of wavelengths within the resonant cavity in each direction must be an exact integer.

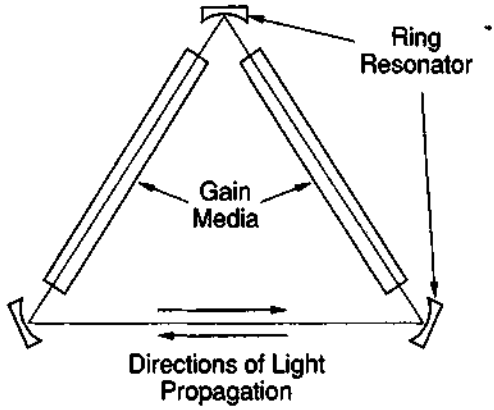


Figure 11-6 A triangular arrangement for a ring laser gyro.

To satisfy this condition, the two beams must have slightly different frequencies. This phenomenon is called the Sagnac effect; it is the basis for the measurement of angular rotation rate.

To understand the operation of the ring laser gyro, it is instructive to work through the mathematical conditions. Consider light leaving point Q in Figure 11-7 and making one transit around the triangle. The point Q will move a small distance during the time that the light takes to traverse the ring. The time T for the light to make one trip around the ring is

$$T = \frac{P}{c} = \frac{3L}{c} \tag{11.5}$$

where P is the perimeter of the triangle, L is the length of one side, and c is the velocity of light. The distance D moved by the point Q in this time is

$$D = \omega T x \tag{11.6}$$

where ω is the angular rotation rate and x is the distance from Q to the intersection of the perpendicular bisectors of the three sides of the equilateral triangle, as illustrated in Figure 11-7. The change in path ΔP is the component of D along the direction of propagation of the light:

$$\Delta P = D \cos 60^\circ = \frac{D}{2} \tag{11.7}$$

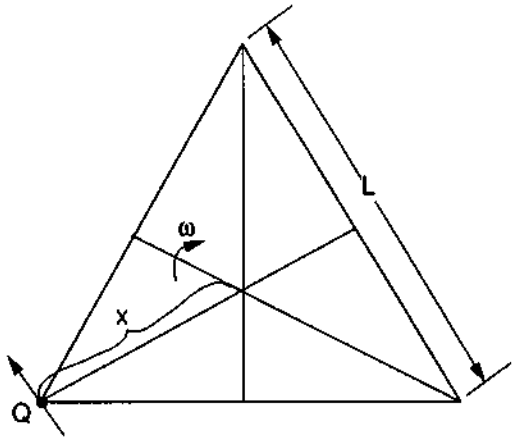


Figure 11-7 Diagram for analysis of ring laser gyro operation. In this figure, ω is the component of angular rotation rate perpendicular to the plane of the triangle.

Using Equation (11.6), we obtain

$$\Delta P = \frac{\omega T x}{2} = \frac{\sqrt{3}\omega L^2}{2c} \tag{11.8}$$

where we have used $x = \sqrt{3}L/3$. Because the area A of the equilateral triangle is

$$A = \frac{\sqrt{3}L2}{4} \tag{11.9}$$

we may rewrite

$$\Delta P = \frac{2\omega A}{c} \tag{11.10}$$

For a standing light wave,

$$N\lambda = P \tag{11.11}$$

where N is an integer; thus, when the path length shifts by ΔP, the wavelength shifts by $\Delta\lambda$, given by

$$\Delta\lambda = \frac{\Delta P}{N} = \frac{\lambda\,\Delta P}{P} \tag{11.12}$$

and the frequency shifts by Δf,

$$\frac{\Delta f}{f} = \frac{\Delta\lambda}{\lambda} = \frac{\Delta P}{P} \tag{11.13}$$

Each of the counterrotating beams will experience the same shift, although in opposite senses. Hence, the difference Δ in frequency between the two counterrotating beams is

$$\Delta = 2\,\Delta f = \frac{2f\,\Delta P}{P} = \frac{4\omega A}{\lambda P} \tag{11.14}$$

This equation is the basic equation that is used for measurement of angular rotation rate. Although it has been derived here for a triangular geometry, it is valid for any path with enclosed area A and perimeter P. It is valid even when general relativity is taken into account. The frequency shift is directly proportional to the angular rotation rate. The frequency shift can be detected as a beat note according to methods described in Chapter 9. For a known geometry and wavelength, measurement of the beat frequency and use of Equation (11.14) allows direct determination of the angular rotation rate.

Specifically, for an equilateral triangle with legs of length L, the equation becomes

$$\Delta = \frac{4\omega HL/2}{\lambda 3L} = \frac{4\omega\sqrt{3}L^2/\theta}{\lambda 3L} = \frac{\sqrt{3}\,\omega L}{6\lambda} \tag{11.15}$$

where $H = \sqrt{3}L/4$ is the height of the triangle. According to this equation, to increase the beat frequency and increase the sensitivity of the measurement, one should increase the size of the triangle and decrease the laser wavelength.

As a specific example, consider a ring laser gyro that uses a helium–neon laser with wavelength 633 nm and a triangle with 10 cm legs. For a rotation rate of 0.1 rad/hr, we obtain a beat frequency of 2.5 Hz. This frequency is easily detectable, although it is only about 2×10^{-14} of the laser frequency.

We note that this measurement yields one component of the angular rotation rate, the component perpendicular to the plane of the triangle. For measurement of all three components, three noncoplanar gyros are required.

For readout of the beat frequency, part of each counterrotating beam is extracted through a semitransparent mirror, and the two beams are combined at a detector or detectors. The beams will have slightly different frequencies, and the frequency difference will be detectable as a beat note in the output of the detector. One method for extracting the beams is shown in Figure 11-8. The two beams emerge with a small angular separation. This leads to formation of a fringe pattern, as indicated in the section A–A. The fringe pattern will be in motion because of the difference in frequencies of the beams. Detectors located in the fringe pattern measure the motion of the fringes from the amplitude modulation of light reaching them. The higher the rotation rate, the greater the frequency difference and the faster the rate of fringe motion.

In principle, ring laser gyros should be very sensitive. In practice, the limits are set by design and construction factors, cleanliness, and by a phenomenon called "lock-in." Lock-in has proved to be a troublesome limitation. It arises because the frequency difference between the two counterrotating beams falls to zero for small

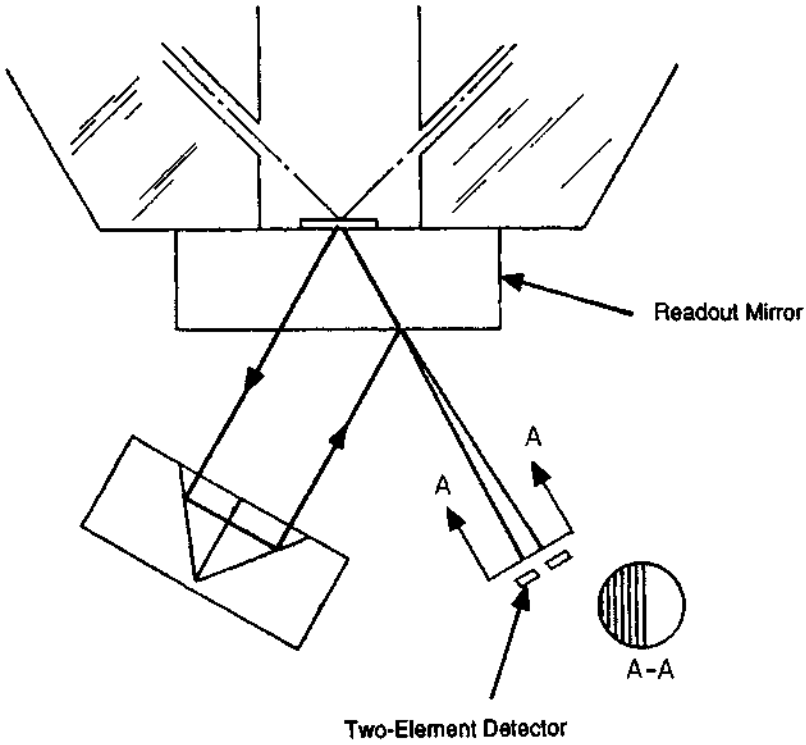


Figure 11-8 Readout technique for ring laser gyro.

values of the angular rotation rate. This is the same as the locking together of the frequencies of two coupled oscillators operating at slightly different frequencies. Thus, there is a threshold value of the angular rotation rate, below which the laser gyro gives no output signal. This is illustrated in Figure 11-9. Various techniques have been devised to reduce the lock-in problem. One successful technique involves dithering the gyro, that is, rotating it back and forth at a high frequency, so that the total rotation rate almost always lies above the lock-in threshold. When one integrates over a number of periods of the oscillatory dither, the contributions of the beat frequencies from opposite portions of the dither motion will cancel. The remaining signal is due to the rotation that is to be measured. With this approach, one can measure rotation rates that would otherwise have been well below the threshold for lock-in.

One of the limiting factors for ring laser gyro performance is the random drift rate, which is one source of noise in the gyro. In a dithered gyro, a small error is produced every time the dither direction reverses and it the rotation rate falls briefly below the lock-in threshold. The random drift rate, measured in degrees per $(\text{hour})^{1/2}$, has been reduced substantially over a period of years, and it can be less than 0.001 $\text{deg}/(\text{hour})^{1/2}$. This value is acceptable for many applications in navigation.

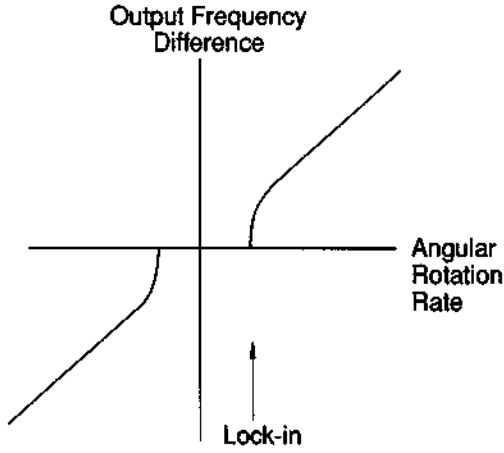


Figure 11-9 Effect of lock-in at low angular rotation rates on the ring laser gyro output.

A fundamental noise limit, called quantum noise, arises from spontaneous emission in the laser cavity. This is also measured in degrees per $(\text{hour})^{1/2}$. The quantum noise decreases with increasing gyro size, but even for very small gyros with 2 in. legs, it is around 0.0003 $\text{deg}/(\text{hour})^{1/2}$. The random drift rates of modern ring laser gyros are approaching the quantum limit.

Construction of ring laser gyros can be relatively simple. The gyro is fabricated from a block of glasslike material with low coefficient of thermal expansion. Holes are drilled to define the legs of the triangular gyro. Electrodes are inserted and mirrors are attached to the block. The structure is filled with a mixture of helium and neon gases and sealed. The resulting structure is rugged and durable and relatively inexpensive to fabricate. A typical example of the structure is illustrated in Figure 11-10. The electric discharge that sustains the laser operation typically fills only a portion of the triangle, as indicated.

A ring laser gyro is an angular rate sensor. To provide the basis of a navigation system, typically three ring laser gyros are orthogonally mounted in an avionics box, along with three accelerometers and a computer. The computer processes the outputs of the sensors and converts the results to data suitable for navigation. Typical accuracy for a medium-performance ring laser gyro navigation system is around one nautical mile of error per hour of flight. In some advanced systems, a hexad of six ring laser gyros, arranged in a noncoplanar configuration, allows additional fault tolerance.

Ring laser gyros have been used as the basis of the navigation systems on many commercial and military aircraft for many years. They use helium–neon as the

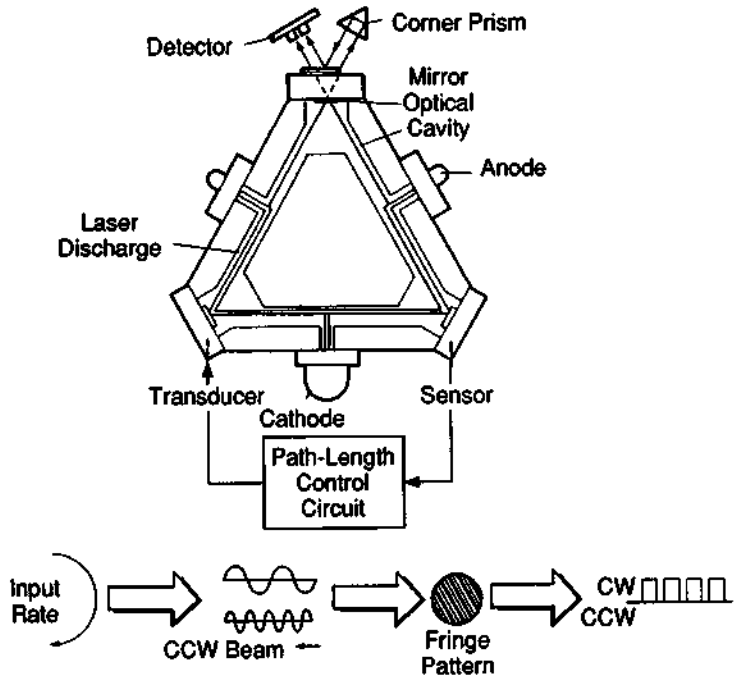


Figure 11-10 Representative structure of a ring laser gyro.

active medium, and most devices operate at a wavelength of 633 nm. Most installed ring laser gyro systems have employed a leg length of 4.2 in., but modern miniature gyros with a leg length as small as 0.8 in. are available.

Compared with conventional rotating gyros, ring laser gyros offer a number of advantages:

- For comparable performance, they are lower in cost.
- They are "strap-down" systems (i.e., rigidly attached to the vehicle) that do not require expensive gimbal mounts.
- They have no high-speed moving parts. This leads to higher reliability.
- They are rugged and durable, with lifetimes of tens of thousands of hours.
- They have very high shock resistance.
- They do not require time to "spin up," as conventional gyros do. They have no heaters and require no warm-up.

To date, the primary application for ring laser gyros has been inertial navigation for aircraft. The ring laser gyro has become the dominant technology for inertial navigation of commercial airliners, and it is widely used in many military and business aircraft. In addition, new applications are developing. These include guidance of many other types of vehicles (such as missiles, munitions, reentry vehicles, helicoptors, self-propelled artillery, submarines, and ships), pointing and tracking for fire control platforms on tanks, control of drilling equipment, pointing and tracking for high-energy lasers, and so forth. The ring laser gyro has provided a novel type of sensor that illustrates dramatically the capabilities of laser technology for measurement applications.

A possible competitor to the ring laser gyro is the fiber optic gyro (FOG). Lengths of optical fiber can be wrapped into coils. Counterrotating light waves in the coils provide a beat frequency in response to angular rotation, just as in the ring laser gyro. This can provide an optical gyro with a large ratio of area to perimeter. According to Equation (11.14), this is a desirable feature. This fact has generated considerable interest in fiber optic gyros, and substantial development work has been expended on them. However, they have not yet reached the status of inertial navigation devices, although some commercial systems for platform stabilization and for measurement of attitude and heading have been developed.

A typical FOG will contain coiled optical fiber, with total length ranging up to 1200 m. According to Equation (11.14), the frequency shift and hence the sensitivity will increase as the length of the fiber increases. The light source is usually an LED or a laser diode operated below threshold. A coupler is used to split the beam and send it in opposite directions through the coil. A detector measures shifts in intensity after the two beams have been recombined by a coupler. The shifts in intensity arise from shifts in the relative phases of the two beams due to rotation of the coil.

Fiber optic gyros are still being developed. They offer a potentially lower-cost and lower-performance alternative to angular rotation sensing. It appears likely that by the late 1990s, they may replace ring laser gyros for applications requiring modest performance, perhaps in the range around 0.01 to 0.1 deg/hr. Applications requiring high performance, down to 0.000015 deg/hr, will continue to be dominated by ring laser gyros.

There are a number of other types of optical fiber sensors. A discussion of these is presented in Chapter 22.

For commercial aircraft navigation, the development of the global positioning satellite (GPS) navigational system may lead to reduced requirements for on-board inertial navigation systems. For the present, regulations still require that airliners have their own on-board navigational equipment. It is possible that in the future there may be less need for such equipment. But even if the role of inertial navigation equipment in commercial aviation were to shrink, the developing uses of ring laser gyro equipment for guidance and navigation of other vehicles indicates continued application of this laser technology for the foreseeable future.

C. Diffractive Measurements of Small Dimension: Wire Diameter

Measurements of the dimensions of a small object can be achieved with high accuracy through observations of the diffraction pattern formed when a laser beam is diffracted around the edges of the object. A specific example of this procedure is the measurement of wire diameter.

The diffraction pattern produced by a wire inserted in the path of a laser beam with diameter larger than the wire is an array of spots extending in a line perpendicular to the wire. Figure 11-11 represents this situation. An undiffracted spot appears in the center of the pattern, which is projected on a screen. The diffracted spots lie

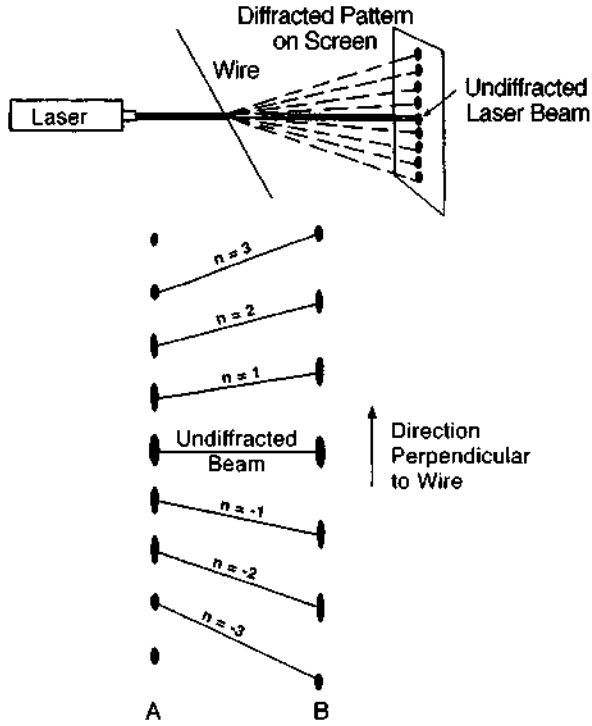


Figure 11-11 Top: Diffractive measurement of wire diameter. Bottom: Change of the diffraction pattern with wire diameter. The wire that produces pattern *A* has a larger diameter than the one that yeilds pattern *B*. In the figure, *n* is the order of the diffraction.

in a line in the direction perpendicular to the wire. The separation of the spots varies inversely with the wire diameter D according to

$$\sin \phi_n = \frac{n\lambda}{D} \tag{11.16}$$

where ϕ_n is the angle of the nth spot from the direction of the laser beam and λ is the laser wavelength. This equation may be used to obtain D from geometrical measurements of the positions where the spots appear. The top portion of the figure shows the pattern for a wire of some diameter. The bottom portion of the figure shows how the pattern changes as the wire diameter decreases: The spots spread farther apart.

This type of measurement has attractive features. There is no contact with the wire. Motion of the wire perpendicular to its length does not affect the separation of the spots, so long as the wire remains within the laser beam. Thus, vibration will not affect the results. Movement of the wire along the direction of its length does not affect the diffraction pattern. Thus, wires can be measured while they are moving, for example, during extrusion. One could obtain a continuous real-time measurement of changes in the wire diameter as it is extruded or drawn. The method is particularly applicable to small wires. In practical terms, this method can yield measurements of diameters of wires or fibers down to 0.00025 cm to an accuracy of 0.5 percent.

This is an application that makes use of the high radiance of the laser. Diffractive phenomena have long been known, but available previous nonlaser sources did not have high enough radiance to make this type of measurement attractive.

An automated system has been developed [2] that uses an array of photodiodes or a vidicon to detect the diffraction pattern. The signal from the detectors is analyzed to determine the separation of the maxima in the diffraction pattern and to give a direct readout of the wire diameter.

Measurements involving diffraction can be used in a variety of ways to determine dimensions of products besides fibers and wires. For example, one of the edges of a part that is to be measured could be used as one side of a narrow aperture. A stationary flat edge would form the other side of the aperture. The part would then be moved in a direction parallel to the length of the aperture. If the dimensions of the product change, the width of the aperture would change. The spatial pattern of light diffracted by the aperture would then change. Photodetectors sensing the change in the pattern could be used to determine the changing dimension of the part. Alternatively, the part could be left stationary and the laser beam scanned along the length of the aperture. Changes in the diffraction pattern as a function of beam position could be used to determine parameters such as flatness or waviness of the surface. Such measurements can be made with very high accuracy and can sensitively detect small changes in the dimension or surface profile of an object. This method can provide solutions to measurement problems for which conventional techniques are difficult to apply.

Some functions that can be performed by diffraction measurements include diameter or width of extruded or drawn products or manufactured parts, thickness of

items produced in sheet form such as plastic, rubber, or paper, flatness or waviness of surfaces of parts such as seals or piston rings, concentricity of manufactured parts, proximity of items such as turbine blades or compressor parts, angular orientation of piece parts, and dimensions of small gaps in products like tape heads.

The diffractive approach to dimensional measurement can yield very high sensitivity to small changes in critical dimensions. It is most applicable for use with items having relatively small dimensions, like wires, or small gaps, like tape heads. Some commercial instrumentation has been developed, aimed at solving a variety of commonly encountered measurement, gauging, and inspection problems. Also, many forms of individualized diffractive instrumentation have been devised by scientists and engineers to solve their own specialized problems. The helium–neon laser has been used most often for this type of application, but visible red semiconductor laser diodes can also be employed.

D. Profile and Surface Position Measurement

Contouring measurements are commonly made with the use of a sharp probe. The probe is in mechanical contact with the surface. Changes in position of the surface induce motion in the probe that can be measured. In this way, a profile of the surface can be mapped out. Such mechanical devices have been used in a variety of applications. Examples include examination of manufactured parts for surface finish or dimensional tolerances, continuous monitoring of thickness or surface parameters in automated process control (with feedback servo control of the machining operation), and accurate determination of the dimensions of models. Profileometry of models in the automotive industry is performed to obtain accurate dimensions for full-scale fabrication.

Use of a light beam as an optical probe in a profileometer offers a number of obvious advantages. There is no physical contact and no possibility of surface damage. Measurements can be made on soft materials, like clay models. Also, there is no wear on the measuring tool.

A variety of techniques for noncontact laser profileometry have been demonstrated, and commercial models have become available. The devices that have been developed can measure a wide variety of materials with surface finishes ranging from polished metals to dull plastics. In many cases it has been possible to obtain higher accuracy than is possible using stylus profileometers.

The basic ingredient of optical profiling systems is the measurement of the distance from the measuring head, located at a fixed position, to some surface. If the measurement is performed with high spatial resolution over the surface, it is basically a profiling operation that generates a contour of the surface. If the measurement is performed with relatively coarse resolution, it can be considered more as a measurement of surface position. It can be performed on a product that is moved continuously under the measuring head. Such a measurement can be useful for determining thickness of the product.

The lasers that have been used for these measurements have usually been helium–neon lasers. Recently, though, visible and infrared semiconductor laser diodes have been used.

Many different approaches to profiling measurement have been demonstrated. We shall present several examples to illustrate the principles that have been used, although these are not an exhaustive list of all laser-based profiling instrumentation.

Triangulation Devices

One commonly employed method for surface profiling utilizes optical triangulation. A spot of light is projected from the measuring head onto the surface. Figure 11-12 shows a simplified diagram of the operation of the device. The source is indicated as a laser diode in this drawing. Light reflected from the surface is imaged onto the detector. The detector may be an array of photodiodes or a position-sensitive lateral-effect photodiode, as described in Chapter 8. If the position of the object surface changes by Δz, the position of the imaged spot is shifted by Δx. The magnitude of the shift in the position of the imaged spot is measured by the detector, and the change in surface position is determined from

$$\Delta z = \frac{\Delta x \sin \phi}{m \sin \theta} \tag{11.17}$$

where m is the magnification of the lens, equal to the distance from the lens to the detector divided by the distance from the surface to the lens, ϕ is the angle between

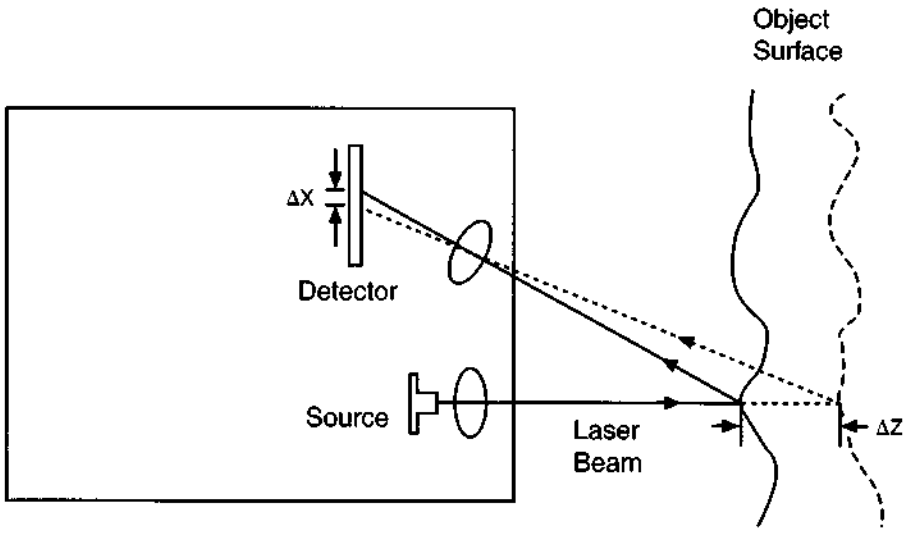


Figure 11-12 Optical triangulation measurement. Δx is the motion of the beam on the detector when the surface moves by Δz from its original position, indicated by the dashed line.

the plane of the detector and the direction of the light striking it, and θ is the angle between the initial light beam and the reflected light. This relation is approximate, valid under conditions where the angles and distances do not change much as a result of the change in position of the surface.

To obtain a contour of the surface, the surface can be translated under the measuring head. Optical triangulation devices have been used for applications such as determining the contours of manufactured parts, measuring the contours of prototypes, and measuring machine tool wear.

Optical triangulation appears to form the basis of the most common commercial laser-based profiling equipment. Commercial models offer resolution down to the micrometer range. The working distance from the head to the surface tends to be small, in the 1 in. range.

Two-Spot System

Another related system uses two laser beams that strike the surface at a known angle, as shown in Figure 11-13. The two beams produce two spots of light on the surface. The separation of the spots depends on the distance from the sensor head to the surface. The spot spacing S is given by

$$S = \frac{2D}{\tan \theta} \tag{11.18}$$

where D is the displacement of the surface from a nominal reference position and θ is the angle between the beam and the plane of the surface. The two spots come together when the surface is at its reference position. As the target surface changes position, the spots move closer together or farther apart, depending on whether the surface position is moving farther from or nearer to the sensor head.

The sensor head contains a lens assembly, an optical scanning system, and a detector that generates two pulses when the positions of the two spots of light are scanned over. The temporal separation of the two pulses is proportional to the spatial separation of the two spots of light. Thus, the time between pulses depends on the position of the target surface. The system is quoted as having the following specifications: accuracy ±0.0003 in., repeatability 0.0001 in., dynamic range 0.2 in., and response speed 1 msec.

Vibrating Pinhole Profile Measurement System

In a more sophisticated system, laser light reflected from the surface being inspected is imaged onto a vibrating pinhole. The apparatus is shown in Figure 11-14. A lens focuses the laser beam onto a small spot on the surface of the object. The same lens collects light scattered from the surface. Another lens then images this light on a pinhole that is driven by a tuning fork oscillator. Light passing through the pinhole reaches a photodetector. The alternating signal from the photodetector is amplified and rectified by a lock-in detector, synchronized with the signal from the tuning fork oscillator. Whether or not the laser beam is focused on the surface can be determined

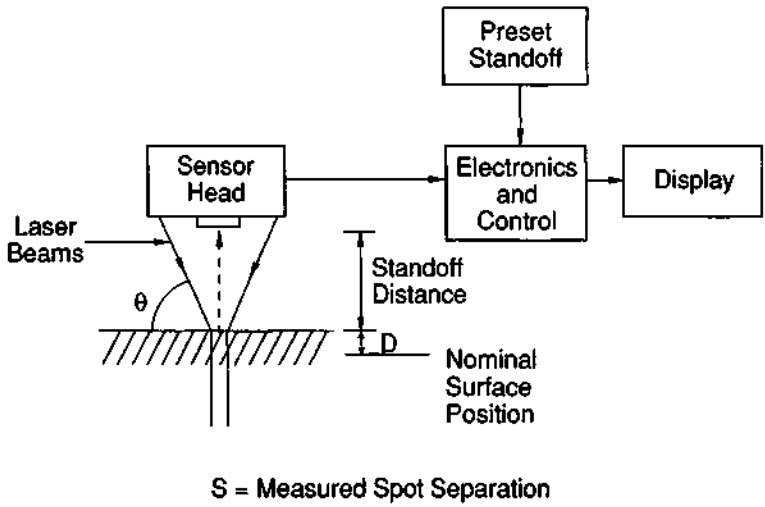


Figure 11-13 System for surface position measurement using two laser beams that strike the target surface at an angle θ. (From D. L. Cullen, reprinted from the *Proceedings of the Technical Program, Electro-Optical Systems Design Conference, 1970*, page 783. Copyright © 1970 by Industrial & Scientific Conference Management, Inc.)

from the output signal of the lock-in detector. The output voltage of this detector, as a function of displacement of the surface, is shown in Figure 11-15. These results were obtained with a 250-μm-diameter pinhole oscillating at 525 Hz. The sign of the output signal corresponds to the direction of displacement of the surface from the focus of the laser beam.

Contour maps can be obtained by moving the object in the two dimensions perpendicular to the incident beam. The reproducibility of the measurements was within 2 μm for surfaces at rest. The accuracy was estimated to be within 40 μm for plaster surfaces and in the 50–80 μm range for rough brass surfaces. The resolution of the system corresponded to the diameter of the focused laser beam, which could be as small as 5 μm.

Such an optical profiling system could be employed for model scanning in an automated system for functions such as die cutting. It could also measure surface roughness on machined parts.

Focus Method

In another variation of laser profiling, the laser light is focused onto the surface by an objective lens. Light reflected by the surface is directed to a detector, which is part of a control loop that adjusts the position of the probe head so that the beam

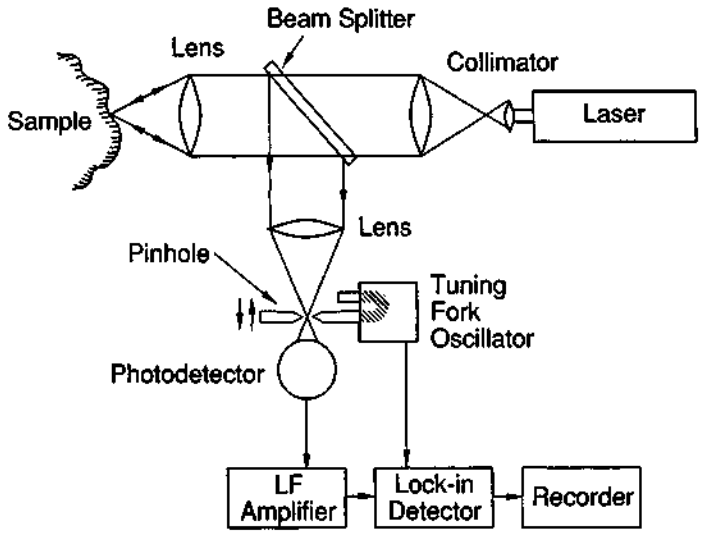


Figure 11-14 Experimental arrangement for profile measurement system. (From S. Ando *et al.*, *Appl. Opt.* **5**, 1961 (1966).)

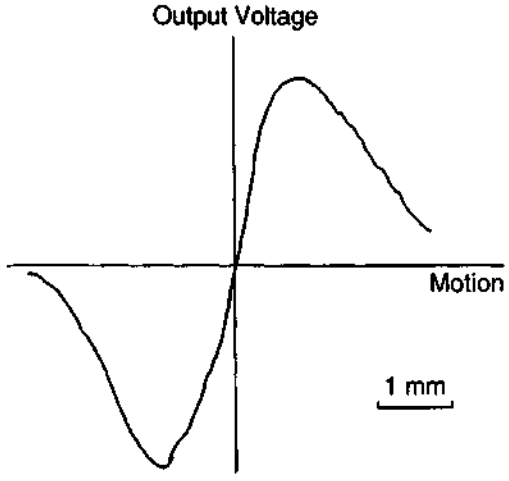


Figure 11-15 Output voltage as a function of surface motion for the profile measurement system. (From S. Ando *et al.*, *Appl. Opt.* **5**, 1961 (1966).)

remains focused on the surface. Thus, the probe stays at a constant distance from the surface. As the probe is moved across the surface, the position of the head follows the surface profile. Inductive coupling is used to convert the motion of the head to an electrical output, which provides a measure of the surface contour. Commercial models of profileometers using this principle offer submicrometer resolution.

Interferometric Systems

Several systems have combined laser interferometry and optical microscopy to produce versatile noncontact profiling devices. A number of approaches have been adopted. In a typical configuration, a helium–neon laser beam is projected onto the test surface through an objective lens. Light reflected from the surface is collected and combined with a portion of the light that is reflected from a reference surface. The two wavefronts form an interference pattern. Because the light reflected from the surface contains information about the surface, the interference pattern also contains that information. Variations in the interference pattern are related to variations in the surface profile. The interference pattern is imaged onto a solid state detector array. The reference surface is moved in a known fashion by a piezoelectric transducer. This changes the relative phase of the light in the two wavefronts, producing a change in the interference pattern. The detector array then views a changing interference pattern. The extra information gained from the movement of the reference surface allows a computer to analyze the interferogram and determine the phase difference between the two wavefronts. From this information, the relative distances that the wavefronts have traveled may be extracted, and the surface profile may be determined. The algorithms employed for the analysis allow reconstruction of the surface height.

In one variation on the interferometric theme, the light from a semiconductor diode laser is delivered to the surface through an optical fiber. The interference occurs between light partially reflected back from the tip of the fiber and light reflected from the surface under test. As the beam is scanned over the surface, variations in the surface height lead to changes in phase between the two interfering beams. The use of the fiber probe leads to very fine lateral resolution in this instrument.

A number of commercial instruments are available. They rely heavily on intensive computer processing and offer many versatile options for display of the surface characteristics. The measurement times are in the range of a few seconds. These interferometric profilers offer vertical resolution down to the subnanometer regime and, typically, horizontal resolution of a few micrometers. But their vertical range is limited, typically no more than a few tens of micrometers. Thus, they are best suited for examination of precise structures with very fine features, like solid state electronic circuitry.

The interferometric profilers will have a different range of applications than the other types of profilers we have described. They have typically been used to measure the surface profiles of objects like diamond-turned mirrors, silicon wafers, optical disks, magnetic disks, and ring laser gyro mirrors. They provide the capability

for very precise evaluation of parameters like surface roughness, step heights of layers in semiconductor structures, and profiles of fine gratings.

The emphasis in our discussion of surface measuring instruments has been on measurement of surface profiles, such as the shapes of models, surface roughness, and so forth. It is sometimes of interest simply to measure the position of a surface. The instrumentation we have described, especially the first three types, can also be used to measure where a surface is. Often the measurement is performed to determine how far the surface is from some reference location. The triangulation method and the two-spot method are well suited to making that type of measurement.

It is also possible to make thickness measurements. These measurements can be made by using two sensors, one to determine the position of each side of an object. The thickness is then determined simply by taking the difference in the positions of the two surfaces. If the object is on a support, such as a conveyer or rollers, the measurement can be made from one side by measuring the position of the support surface and then measuring the distance of the object surface from that reference position. We shall discuss this type of measurement in more detail in the next section.

E. Measurement of Product Dimension

Lasers have been used in a variety of ways for measuring the dimensions of manufactured parts or products. The discussion in the previous section briefly mentioned some of the possible approaches. This section will summarize some of the available procedures for measuring product dimension and compare the methods.

Use of laser light for measuring product dimensions has taken a wide variety of forms. Because of the enormous diversity of manufactured products and the variety of requirements for gauging them, it is not possible to describe a generalized system. Rather, individual systems have been devised to meet the specialized nature of specific applications.

Lasers have been used for measuring a wide range of products, including sheet thickness of plastics, extruded rubber, and metals, dimensions of piston rings, cylinder heads, spark plugs, fuel injector components, and choke springs in the automotive industry, the diameter of glass tubing, fuel rods, and munitions, the shape of items such as roller bearings and valve lifters, concentricity and wall thickness for tubes and hoses, the thickness of moving webs of products like cloth and paper, the amount of material used in food products, the length of pieces of lumber, and the fill level of plastic pouches containing powdered products. This is not a complete list; it is merely an indication of the diversity of the capabilities of laser measuring systems. The specific method by which a laser system can be applied to a particular product necessarily depends on the measuring procedure and on the requirements for the specific product.

There have been three main techniques by which dimensional measurements have been made:

1. Measurement of the obscuration of a laser beam that passes across the item

2. Dimensional comparison measurements in which the position of the surface of the product is determined relative to the sensor head, such as the triangulation method discussed in the last section
3. Diffraction-type measurements, which have been described earlier in relation to measurements of wire and fiber diameters.

The first of these types of measurement involves interposing the object to be measured in the path of a beam that is scanned across the object. The beam is detected by a photodetector. An object in the path of the beam interrupts the detector output for a time. In one arrangement, the interruption produces a timing signal. A digital readout that is generated from the timing signals as the beam crosses the edges of the object provides a direct display of the dimension. An alternative method uses a stationary laser beam through which the objects pass, as on a conveyer belt.

Devices using this approach can measure fast-moving parts or continuously produced products without precise positioning. Measurements may be made in one dimension or in two perpendicular dimensions. The measurement results can be used for closed-loop control of the manufacturing process. Small modifications to the basic measurement technique can be implemented to measure profiles (by moving the scan along the length of the object) or roundness (by rotating the object in the beam).

This method is the simplest to implement and the easiest to interpret. It requires calibration because of the nonuniform spatial profile of the laser beam. The edge of the beam is not infinitely sharp, so that the detector signal decreases as a function of time as the beam sweeps across the edge. Thus, a calibration is needed to determine the proper signal level to accept as the position of the edge.

The second method, dimensional comparison, makes use of surface position measuring equipment like the triangulation systems and two-spot systems described in the last section. The position of a surface of the object is measured in relation to the position of the sensor. With two sensors, each measuring the position of a different surface, one can measure the thickness of a sheet of material. Alternatively, if parts are moving on a belt, the position of the belt can serve as a zero reference. Then, as parts move under the beam, the thickness is determined by comparison of the position of the top of the object and the reference surface.

The diffractive type of measurement offers great versatility and the highest resolution. It is not applicable to all products. A system must be devised in which a small dimension, such as a slit, can be measured. Large parts are less suitable than small parts, and the total range over which measurements can be made is limited.

Some characteristics of commercial models of each of these three approaches are presented in Table 11-1. The scanned beam and dimensional comparison methods offer comparable resolution and accuracy, with the scanned beam method offering larger ranges at which measurements can be made. The diffractive method offers the highest resolution and best accuracy, but it is somewhat limited in its applications.

Although the specific implementation of dimensional measurement for process control or inspection for a particular part or product varies according to the shape,

Table 11-1 **Typical Characteristics of Laser Gauging Systems**

Measurement type	Minimum resolution	Accuracy	Maximum range of measurement	Typical use
Scanned beam	0.5–5 μm	0.01 to 0.1% of measurement	Up to feet	Diameter of rods, profiles of extruded parts
Dimensional comparison (triangulation)	2–10 μm	0.1% of measurement	To 50 cm	Thickness of sheet products
Diffraction	0.025 μm	0.12 μm	0.25 cm	Wire diameter, width of small gaps

size, and requirements of an individual application, the general principles described in this section can be used for a broad range of automated inspection processes in an industrial setting. Choice of a method will depend on the specific requirements of the application.

Laser gauging as described here offers the advantage of being a remote noncontact gauging process, with capability for high speed, reproducibility, and good accuracy.

F. Measurement of Surface Finish

The profiling systems described in Section 11.D can be used to inspect surface flatness and roughness, at least to within the tolerances set by their specifications. In this section, we shall describe alternative methods of surface inspection that aim more at detection of imperfections on a surface. These approaches rely on the scattering of light by surface imperfections.

Various surface defects, such as scratches, pits, and small raised areas, reflect light differently than perfect surface areas. Scattered light from the defects can be detected simply by a photodetector. The technique is complicated by background reflection and scattering from other areas on the surface. In practice, it is difficult to distinguish between light scattered from defects and light scattered in all directions by the general roughness of the surface.

One technique for resolving this problem has used illumination of the surface at a grazing angle coupled with spatial filtering to eliminate the specular component of reflected light. A typical arrangement is shown in Figure 11-16. The spatial filter contains a central stop that passes only scattered light and eliminates the central diffraction peak due to specularly scattered light.

Some experimental results are shown in Figure 11-17 for a steel plate with a surface finish of 0.2 μm. Scratches about 10 μm wide and 5 mm long were made on the plate with a diamond scriber. The scratches were perpendicular to the plane of incidence of the light, but the scattering was not a strong function of orientation. The results are shown for an angle of incidence of 88°, which maximizes the signal

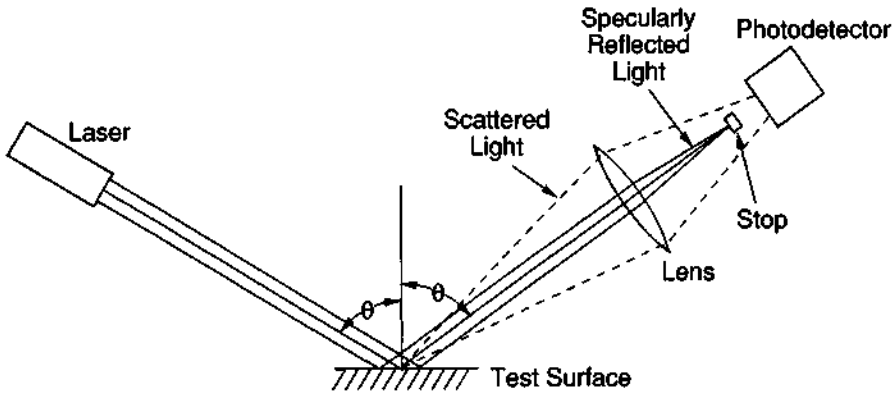


Figure 11-16 Arrangement for measurement of surface finish. (From T. Sawatari, *Appl. Opt.* **11**, 1337 (1972).)

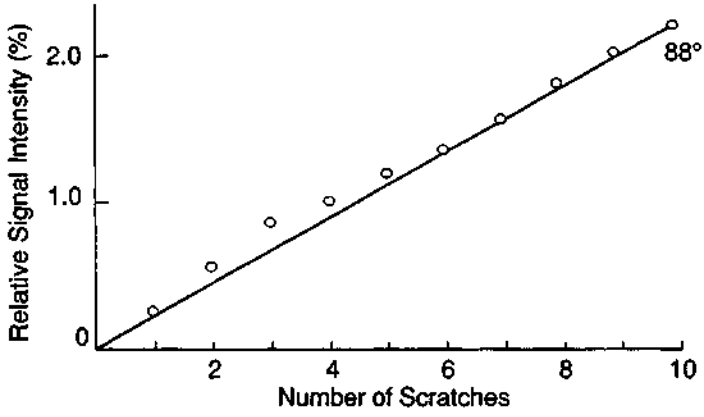


Figure 11-17 Experimental results for measurement of surface finish. (From T. Sawatari, *Appl. Opt.* **11**, 1337 (1972).)

under these conditions. The system clearly distinguishes between a lightly scratched and a heavily scratched surface.

An alternative arrangement uses a scanning helium–neon laser beam to detect the location of individual scratches. An oscillating mirror sweeps the beam back and forth across the surface, which moves in a direction perpendicular to the scan. The intensity of the scattered light increases as the beam passes over a defect. A photodetector detects the scattered light. The output of the detector is used to map the number, relative size, and location of defects. The map may be generated and displayed automatically as the beam scans the surface.

A system using this approach has been used to map defects on ring laser gyro mirrors. The resulting map of defects is used both as an acceptance criterion for the mirrors and also to position the mirror so that the region with minimum defect density is in the path of the laser beam in the gyro.

Most surface roughness measuring systems have employed helium–neon lasers, but commercial surface roughness monitors using visible semiconductor laser diodes operating at 670 nm have become available.

These examples show how lasers can be used to inspect surfaces for defects. In applications where high surface perfection is important, laser systems offer a considerable advantage in cost and speed compared with visual inspection under a microscope.

G. Particle Diameter Measurement

There are many requirements for measuring particle size distributions. The particles may be entrained in flowing gases or liquids or may be in the form of dry powders. The environments may be hot or corrosive. The required measurements include diverse materials, such as aerosols, combustion products, reactive powders, particles in paper pulp slurry, and components in industrial process streams.

Laser-based methods offer many attractive features for such measurements. The measurement process is remote and noncontact. It can very rapidly determine the complete particle size distribution.

Several different laser approaches to measurement of particle size distributions have been developed. Perhaps the most common has been scattering of a beam of laser light by the particles. Under conditions where the particle sizes are larger than the laser wavelength, the scattering is relatively simple. The angle of scattering increases as the size of the particles decreases. The beam from a laser, typically a low-power helium–neon laser or a semiconductor diode laser, is expanded and transmitted through the sample containing the particles. The scattered light is collected by a lens and focused onto an array of detectors. Light that was not scattered is focused on the central axis of the system. Light scattered at small angles is focused off-axis, with the distance from the axis increasing as the scattering angle increases. The relative signals from detectors at different distances are analyzed and processed to yield the particle size distribution.

This method is useful for particles with sizes in the range from a few to hundreds of micrometers. For smaller particle sizes, comparable with or smaller than the laser wavelength, the scattering function becomes complicated. For particles larger than a few hundred micrometers, the scattering angle becomes very small, and the measurement is difficult. Other methods have been developed to cover those cases.

Commercial laser-based particle sizing instrumentation is available in many different forms. The devices are based on either helium–neon lasers or semiconductor diode lasers, but most new instruments now employ semiconductor lasers in order to reduce the size of the instrument. The instruments employ a variety of different technical approaches, intended to provide varying capabilities suited to different measurement requirements. Some use multiple lasers traversing the sample at different angles in order to define the complicated scattering patterns characteristic of small particles. To cover a wide range of particle sizes, an instrument may transition from one measuring technique to a different technique. Devices are available covering a variety of different particle size ranges. The user may choose an instrument with a range that covers the particle sizes for a particular application. It is possible to obtain particle size distributions covering the range from 0.04 to 2000 μm in a single scan, within a time of a few tens of seconds.

An example of a configuration that uses three semiconductor laser diodes is shown in Figure 11-18. Light scattered from one beam, designated the main laser, is detected by two different linear detector arrays, one with the detectors at small angles to the beam and one at an angle around 60°. Scattering from the second laser beam (L_2) is detected by the same array of detectors, but at scattering angles greater than 60°. Backscattering from the third laser beam (L_3) is also detected by the same array. This arrangement thus increases the number of detectors available and the angular range covered. The instrument covers a size range from 0.1 to 700 μm, for wet or dry particles.

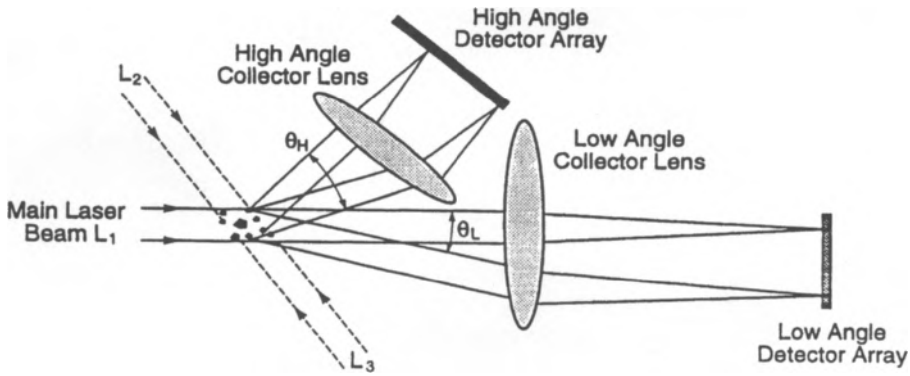


Figure 11-18 A three-laser particle sizing system. The three beams are denoted L_1, L_2, and L_3. (From P. J. Freud, A. H. Clark, and H. N. Frock, at the Pittsburgh Conference, Atlanta, GA, March 8–12, 1993.)

Figure 11-19 shows the performance of the system in measuring size distributions of submicrometer polystyrene spheres. The figure shows the distributions obtained for spheres of diameter 0.13 μm, 0.4 μm, and for the two sizes mixed together.

Such instrumentation is widely used in a variety of manufacturing applications like control of production of polymers, abrasives, inks, tires, food products, and many other items.

H. Strain Measurement

Laser instrumentation to measure surface strain has frequently used measurements on the speckle pattern produced when laser light is reflected from a diffuse surface. The phenomenon of the speckle pattern has been described earlier in this chapter. When an illuminated surface is strained, there will be some motion of the speckle pattern. A variety of different approaches have been developed to derive the surface strain from the change in the speckle pattern.

In one manifestation, a beam from a helium–neon laser is split into two parts and the two resulting beams are directed to the same spot on the test surface, at equal angles from the surface normal and on opposite sides of the normal. The two beams are switched on and off sequentially. The speckle patterns from the two beams are recorded by a linear array of photodiodes, which is aligned in the plane of incidence defined by the two beams and the surface normal. Then, the sample is strained, and the changed speckle patterns are also recorded. This pair of speckle patterns is displaced from the first pair. The displacement depends on the strain of the surface and

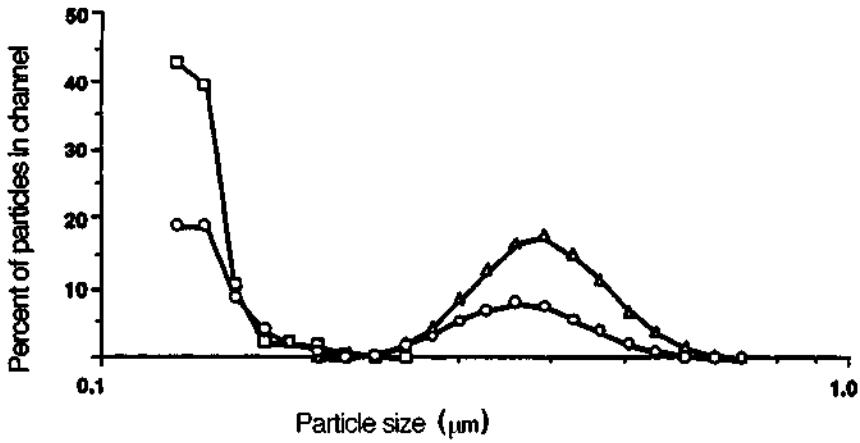


Figure 11-19 Particle size distributions for polystyrene spheres. Squares: 0.13 μm spheres. Triangles: 0.4 μm spheres. Circles: Mixed sizes. (From P. J. Freud, A. H. Clark, and H. N. Frock, at the Pittsburgh Conference, Atlanta, GA, March 8–12, 1993.)

on any rigid-body translation of the sample that may have occurred. The availability of two pairs of speckle patterns allows the effect of the rigid-body translation to be subtracted. The procedure yields one component of surface strain, the component parallel to the plane of incidence. If the apparatus is rotated about the surface normal and measurements are performed with three different planes of incidence, it is possible to derive the two-dimensional surface strain.

Commercial instrumentation based on speckle pattern effects offers resolution as small as 0.3 μm for the displacements resulting from strain.

I. Vibration

Laser measurement of vibrations can be attractive because it is a noncontact method. It will offer advantages for measurements of the vibration of surfaces that are hot, corrosive, or irregular, and for structures that are lightweight or delicate.

Perhaps the most common approach to laser-based vibration measurement has been to split the beam from a low-power laser into two parts. One beam is directed to an acoustooptic modulator, which shifts the frequency of the light slightly. The second beam is directed to the test surface. Light reflected from the surface is collected by a lens and recombined with the first beam. A detector views the combined beams. If the test surface is stationary, the output of the detector will contain a steady beat note at the frequency of the shift imposed by the acoustooptic modulator. If the surface is vibrating, the light reflected from the surface will have its frequency Doppler shifted. The beat note will then be frequency modulated, by an amount proportional to the velocity of the surface. The control electronics process the detector signal, recover the modulation, and convert it to surface velocity. Electronic differentiation of the velocity can yield acceleration.

Commercial laser vibrometers based on this method are marketed. Some use helium–neon lasers, some argon lasers, and some semiconductor diode lasers. Quoted specifications include velocity resolution less than 1 μm/sec and measurements at vibration frequencies ranging from less than 1 Hz to more than 1 MHz.

J. Cylindrical Form Measurement

A relatively recent development involves the capability of measuring the roundness and cylindricity of cylindrical components by means of laser interferometry. The technique employs grazing incidence interferometry. We first define grazing incidence interferometry as it might be used with a flat surface. A right-angle prism is placed adjacent to the surface to be measured, with the side of the prism that represents the hypotenuse next to the surface. The probing beam enters one of the shorter sides, at a large angle θ (perhaps 85°) to the normal to the surface. Part of the light is reflected by the hypotenuse of the prism, and part by the surface being probed. The two reflected components are allowed to interfere. The fringes in the interference pattern represent lines of constant elevation of the surface. The use of the

grazing angle increases the sensitivity of the measurement by a factor of $1/\cos\theta$ compared with the use of conventional Michelson interferometry, in which the beam is perpendicular to the test surface.

This description shows how grazing angle interferometry could be used with a plane surface. To extend it to a cylinder would require the use of a very large number of very thin prisms surrounding the cylindrical form, an impractical configuration. This difficulty can be removed by the use of diffractive optics. Diffractive optics employs the physical phenomenon of diffraction to manipulate optical beams, rather than the phenomena of refraction or reflection, which have been conventionally employed in optical systems. Diffractive optics is a rapidly advancing technology that is being accepted in many optical designs, replacing refractive and reflective optical elements. In many cases, the use of diffractive optics offers advantages in size, weight, and performance compared with conventional optics. Industrial applications of diffractive optics are growing and include use in head-mounted displays, in optical interconnects between chips on a printed circuit board, and in coupling light from laser diodes into optical fibers.

In the present application, the use of a diffractive optical element can produce the two beams required to probe the surface of a cylindrical component. One beam is reflected from the surface of the cylinder, all around the periphery of the cylinder. This wave is distorted by any deviations of the form from a true cylinder. This beam then interferes with an undistorted beam to form a fringe pattern that may be analyzed to determine departures from the desired shape.

This method has been applied to the measurement of the form of diesel engine fuel injectors [3]. These components must meet stringent requirements on the form of the cylinder, and they must be tested rapidly so as not to increase costs excessively. The measurements used a helium–neon laser and diffractive optical elements. The measurement procedures determined the roundness, straightness, and taper of the cylindrical form. The roundness could be determined to within a fraction of a micrometer all around the periphery of the cylinder. The measurements were rapid, in conformance with production needs. The technique may be extended to the measurement of other cylindrical manufactured forms.

K. Defect Detection

Laser instrumentation has found many applications for defect detection. This is basically not a new topic; rather, it is a variation in the utilization of some of the methods already discussed. For example, the scanning beam type of measurement discussed for dimensional measurement may readily be adapted to detection of holes, tears, or other defects in sheet products.

Usually, a helium–neon laser with a few milliwatt output has been used for such applications, but visible semiconductor diode lasers are also used. The beam may be spread into a line or scanned back and forth across materials such as plastic, paper, textiles, or sheet metal. The material moves at high speed perpendicular to the line of light. The presence of rips or tears will lead to an increase in transmitted light,

which may be detected by photodetectors on the opposite side of the sheet from the laser.

Lasers have been used effectively for detecting small tears in the edges of products like paper. A low-power laser beam is incident on one side of the edge of the paper. If an edge tear is present, light reaching a detector on the opposite side will increase. Detection of small edge tears can allow for corrective action before the paper tears completely and jams the machinery.

This is only one example out of many possible laser applications in defect detection.

L. Surface Flaw Inspection Monitor: A Specific Example

As a specific example, we shall discuss an optical surface flaw inspection system that has been developed for 100 percent inspection of small caliber ammunition [4]. This example is chosen because it demonstrates the versatility of laser inspection in a situation in which automatic inspection by conventional techniques had been difficult. The technique could be applied to inspection of surface quality for other mass-produced parts when the surface must be inspected rapidly. The technique could be used in applications like inspection of continuous webs or small precise devices, like automotive parts.

The surface flaw monitor uses scattered light and electrooptical instrumentation to detect the presence of flaws on 100 percent of the cartridge case surface at throughput rates exceeding 1200 cases/min. The inspection of surfaces of cartridge cases had been difficult to implement with automatic techniques because the surface has several discontinuities in profile, a rough surface finish, and acceptable variations in surface coloration. A number of different types of flaws must be detected, including dents, scratches, scaly metal, split cases, folds, wrinkles, buckles, and variations in mouth and primer pocket shape.

The complex case surface and the high speed requirements make optical inspection desirable. There are many types of surface flaws, but all can be determined from deviations from a normal configuration of reflected light. A helium–neon laser beam is spread into a line with cylindrical optics and automatically tracks a spinning cartridge on the perimeter of a wheel-type mechanical handling system. A number of fiber optic elements located at different case positions collect scattered light from surface zones of the case and deliver the light to photomultiplier detectors. The detected signals are processed to determine the frequencies present. The pattern of frequencies in the scattered light relate directly to the type of surface flaw. The signals are fed to a minicomputer, which controls acceptance or rejection by mechanical handling hardware.

At each test station, a spindle holds and spins the cartridge case. As the case turns, the entire surface is illuminated. Scattered light is collected by the fiber optic bundle, which is located off-axis from specularly reflected light. The frequency spectrum is compared with the signature of an acceptable case. The signal-to-background ratio for flaws is large, allowing simple threshold-level signal processing to produce a digital accept or reject signal. The reject signal initiates a reject sequence

that removes the cartridge from the line. As mentioned, the system can achieve a throughput rate of 1200 cartridge cases/min.

M. Summary

Lasers have many possible applications in measurement and inspection. We have described some of them in this chapter. This discussion is not a complete enumeration; rather, it is a collection of examples that may stimulate the imagination of the reader to apply lasers to his or her own particular measurement problem. We will further discuss measurement of strain and vibration in our exploration of the applications of holography.

The phenomena used for measurement have been well known, typically involving diffraction, interference, and so forth. The basic methods are not new. What is new is the light source. The laser has collimation, radiance, and coherence properties that make many types of optical measurement much easier than with conventional light sources. In particular, the field of interferometry has been revitalized by the laser. In many cases, laser-based methods can make certain types of measurement more easily and economically than any other method.

Lasers have been used for many specialized engineering measurements that were previously difficult. These measurements have used instrumentation specially constructed for a particular task and have often been performed by scientists or engineers to solve difficult problems. In addition, many types of commercial laser-based measuring equipment are available and can be used by less skilled personnel on a routine basis. We may expect these types of measurement apparatus to increase as more commercial instrumentation continues to develop.

References

[1] G. Stavis, *Instrum. and Contr. Syst.*, p. 99 (February 1966).
[2] F. P. Gagliano, R. M. Lumley, and L. S. Watkins, *Proc. IEEE* **57**, 114 (1969).
[3] L. Denes and J. Salsbury, at the 1995 Topical Conference on Microscopic Inspection of Macroscopic Structures, American Society of Precision Engineering, 1995.
[4] F. Reich and W. J. Coleman, *Opt. Eng.* **15**, 48 (1976).

Selected Additional References

A. Velocity Measurement

J. B. Abbiss, T. W. Chubb, and E. R. Pike, Laser Doppler Anemometry, *Opt. Laser Technol.* **6**, 249 (1974).
J. W. Bilbro, Atmospheric Laser Doppler Velocimetry: An Overview, *Opt. Eng.* **19**, 533 (1980).
F. Durst, A. Melling, and J. H. Whitelaw, *Principles and Practice of Laser Doppler Anemometry*, Academic Press, New York, 1976.
J. Turner and S. Fraser, *Laser Anenometry*, Springer-Verlag, New York, 1988.

B. Angular Rotation Rate

J. Killpatrick, New Developments in the Ring Laser Gyro, *Scientific Honeyweller*, p. 59 (Fall 1987).

C. Diffractive Measurements of Small Dimension: Wire Diameter

M. Koedam, Determination of Small Dimensions by Diffraction of a Laser Beam, *Philips Tech. Rev.* **27**, 208 (1966).

D. Profile and Surface Position Measurement

D. L. Cullen, Optical Thickness Measurements, *Electro-Optical Systems Design*, p. 52 (July 1975).
T. Kohno *et al.*, High Precision Optical Surface Sensor, *Appl. Opt.* **27**, 103 (1988).

E. Measurement of Product Dimension

Laser Sensors for Accurate Measurement on Moving Webs, *Sensors*, p. 34 (May 1992).
D. M. Papurt and D. A. Cohen, Laser-Based Outside Diameter Measurement, *Quality*, p. 56 (March 1988).
T. R. Pryor and J. L. Cruz, Laser-Based Gauging and Inspection, *Electro-Optical Systems Design*, p. 26 (May 1975).

F. Measurement of Surface Finish

S. S. Charschan, ed., *Lasers in Industry*, Van Nostrand Reinhold, New York, 1972, pp. 335–340.

G. Particle Diameter Measurement

H. G. Barth, ed., *Modern Methods of Particle Size Analysis*, Wiley, New York, 1984.

Chapter 12 | Interaction of High-Power Laser Radiation with Materials

The ability of lasers to produce intense pulses of light energy leads to many applications involving heating, melting, and vaporization. The feature of the laser that allows it to be used in metalworking is, of course, its ability to deliver very high values of irradiance to a workpiece. (Irradiance is defined as the incident laser power per unit area at the surface; it has units of W/cm^2.) A conventional thermal source, such as a welding torch, delivers much lower irradiance, and it cannot be localized so well as a laser. Only an electron beam can compare with a laser in this respect. The total power is not necessarily so important as the ability to focus to a small spot, producing a high irradiance. A 200-W light bulb cannot melt metal; a 200-W continuous laser can do so.

In this chapter, we shall discuss the physical processes that occur during the interaction of high-power laser radiation with materials. Knowledge of these processes is important for understanding the capabilities and limitations of laser-based material processing. These interactions are the basis for laser applications in material processing. We shall emphasize metallic targets, but much of what we say is applicable to other types of absorbing materials.

When laser radiation strikes a target surface, part of it is absorbed and part is reflected. The energy that is absorbed begins to heat the surface. There are several regimes of parameters that should be considered, depending on the time scale and on the irradiance. For example, losses due to thermal conduction are small if the pulse duration is very short, but they can be important for longer pulses. Under some conditions, there can be important effects due to absorption of energy in the plasma formed by vaporized material above the target surface. We note that losses due to thermal reradiation from the target surface are usually insignificant.

The heating effects due to absorption of high-power beams can occur very rapidly. The surface quickly rises to its melting temperature. Laser-induced melting is of interest because of welding applications. One often desires maximum melting under conditions where surface vaporization does not occur. Melting without vaporization is produced only within a fairly narrow range of laser parameters. If the laser

irradiance is too high, the surface begins to vaporize before a significant depth of molten material is produced. This means that there is a maximum irradiance suitable for welding applications. Alternatively, for a given total energy in the laser pulse, it is often desirable to stretch the pulse length.

Melting of a material by laser radiation depends on heat flow in the material. Heat flow depends on the thermal conductivity K. But thermal conductivity is not the only factor that influences the heat flow. The rate of change of temperature also depends on the specific heat c of the material. In fact, the heating rate is inversely proportional to the specific heat per unit volume, which is equal to ρc, where ρ is the material density. The important factor for heat flow is $K/\rho c$. This factor has the dimensions of cm^2/sec, characteristic of a diffusion coefficient. Thus, it is known by the descriptive term "thermal diffusivity," to recognize that it represents the diffusion coefficient for heat.

The factor $K/\rho c$ is involved in all unsteady-state heat flow processes, such as pulsed laser heating. The significance of this material property is that it determines how fast a material will accept and conduct thermal energy. Thus, for welding, high thermal conductivity allows larger penetration of the fusion front with no thermal shock or cracking. Table 12-1 lists the thermal diffusivity of several metals and alloys. The thermal diffusivity of an alloy is usually lower than that of the pure metal that is the major constituent of the alloy. Stainless steel and some nickel alloys have especially low values. A low value of thermal diffusivity for a material limits the penetration of heat into the material and may reduce the laser weldability.

The depth of penetration of heat in time t is given approximately by the equation

$$D = (4kt)^{1/2} \tag{12.1}$$

where D is the depth of penetration of the heat and k is the thermal diffusivity. Let us consider an example. For a metal with thermal diffusivity 0.25 cm^2/sec, heat can penetrate only about 3×10^{-4} cm during a pulse of 90 nsec duration (typical of a Q-switched laser). During a pulse of 100 μsec duration (typical of a normal pulse laser) heat can penetrate about 0.01 cm into the same metal.

These ideas lead to the concept of a thermal time constant for a metal plate of thickness x. The thermal time constant is equal to $x^2/4k$. The thermal time constant represents the length of time it takes for heat to penetrate to a specified depth. Strictly, it is the time required for the temperature increase at the back surface of a plate to reach to 37 percent of the temperature increase at the front surface when heat is absorbed in a very short pulse at the front surface. The thermal time constant gives a convenient order-of-magnitude estimate of the time required for heat flow through the plate.

For effective melting, the laser pulse duration should be approximately equal to the thermal time constant for the metallic sample. Table 12-1 shows values of thermal time constant for some metals. For thin material (0.01–0.02 cm), the thermal time constant is shorter than (or at least comparable with) pulses from normal pulse lasers (millisecond durations). The thermal time constants are much longer than pulse durations characteristic of Q-switched lasers (around 10^{-7} sec). This means

Table 12-1 **Thermal Diffusivity and Thermal Time Constants**

Metal	Thermal diffusivity (cm^2/sec)	Thermal time constants (msec)			
		0.01 cm thick	0.02 cm thick	0.05 cm thick	0.1 cm thick
Silver	1.70	0.015	0.059	0.368	1.47
Aluminum alloys					
Commercially pure	0.850	0.029	0.118	0.74	2.94
2024 alloy	0.706	0.035	0.142	0.89	3.54
A13 Casting alloy	0.474	0.053	0.211	1.32	5.27
Copper alloys					
Electrolytic (99.95%)	1.14	0.022	0.088	0.55	2.19
Cartridge brass	0.378	0.066	0.265	1.65	6.61
Phosphor bronze	0.213	0.117	0.470	2.93	11.74
Iron alloys					
Commercially pure	0.202	0.124	0.495	3.09	12.38
303 stainless steel	0.056	0.446	1.786	11.16	44.64
Carbon steel (1.22 C. 0.35 Mn)	0.119	0.210	0.840	5.25	21.01
Nickel alloys					
Commercially pure	0.220	0.114	0.454	2.84	11.36
Monel	0.055	0.455	1.818	11.36	45.46
Inconel	0.039	0.641	2.564	16.03	64.10

that effective melting through a workpiece is not practical with Q-switched lasers. Lasers suitable for welding are either continuous (e.g., CO_2 or Nd:YAG) or have pulse durations in the millisecond regime (like Nd:glass or ruby).

For somewhat thicker metals (around 0.1 cm), the thermal time constant becomes too long for heat to penetrate through metals with low thermal diffusivity within a pulse of reasonable length. But for metals with high values of thermal diffusivity (like silver, copper, or pure aluminum), the thermal time constant can be a few milliseconds, and heat can penetrate through them during a normal laser pulse.

For even thicker metals (>0.1 cm thick), the thermal time constant becomes long, of the order of milliseconds, even for metals with the highest values of thermal diffusivity. The thermal time constants are much longer than the longest pulse durations that are available. This means that it becomes difficult to weld metals with thickness 0.1 cm or greater. One could, of course, allow a continuous laser to dwell on a spot for a long time to achieve deep penetration, but this approach is not commonly used because the heat spreads over too large an area and the welding rate becomes very low.

Low values of thermal diffusivity mean that the heat does not penetrate well into the material. But a high value of thermal diffusivity can also cause problems by

allowing rapid removal of heat from the surface. This can reduce the amount of melting. To compensate for these effects, one should vary the laser parameters for optimum welding of different metals. To weld copper, for example, one should use a higher power and shorter pulse length to overcome losses due to high thermal diffusivity. To weld stainless steel, one should use lower power and longer pulse duration to achieve good penetration.

Effective melting and welding with lasers depends on propagation of a fusion front through the sample during the time of the interaction, at the same time avoiding vaporization of the surface. Figure 12-1 shows the time dependence of the penetration of the molten front into a massive nickel sample for an absorbed irradiance of 10^5 W/cm^2. About 4 msec after the start of the pulse, the surface begins to vaporize. We note that the depth of penetration without surface vaporization is limited. To obtain greater depth, one can tailor the laser parameters to some extent. Generally, one lowers the irradiance and increases the pulse duration. The control is rather sensitive. One must make careful adjustments to achieve a balance between optimum penetration depth and avoidance of surface vaporization. (The results shown in Figures 12-1 and 12-2 are calculated using an analog computer routine developed by M. I. Cohen [1].)

One is interested primarily in welding under conditions where surface vaporization does not occur. Melting without vaporization is produced only within a narrow range of laser parameters. If the laser irradiance is too high, the surface begins to

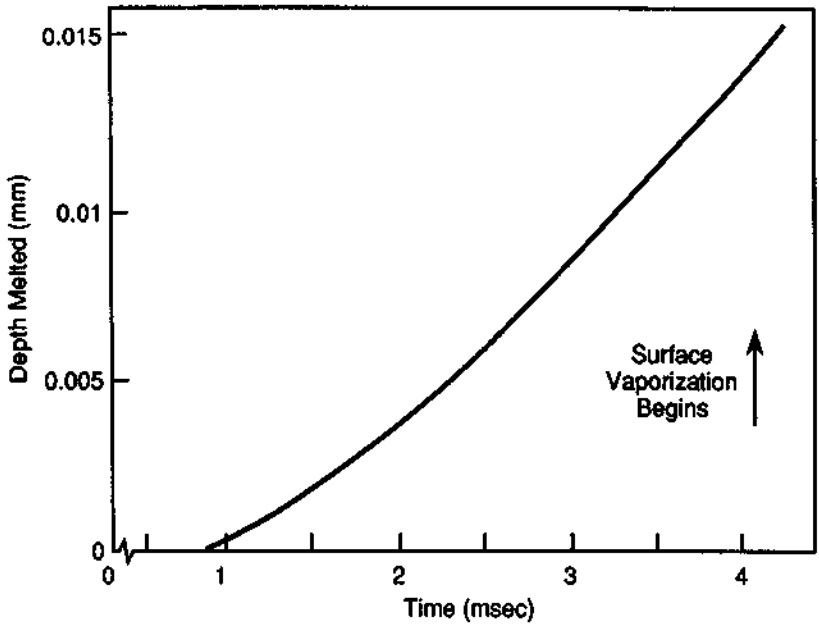


Figure 12-1 Calculated depth melted in nickel as a function of time for an absorbed laser irradiance of 10^5 W/cm^2.

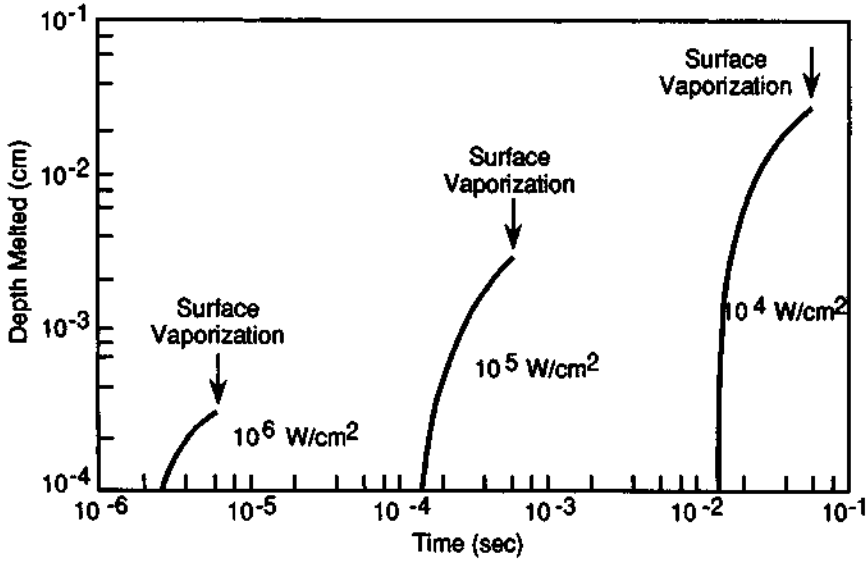


Figure 12-2 Calculated depth melted in stainless steel for several different values of laser irradiance. The time at which surface vaporization begins is indicated for each curve.

vaporize before the fusion front penetrates deep into the material. This means that there is a maximum irradiance suitable for welding applications. Alternatively, for a given total energy in the laser pulse, it is often desirable to stretch the pulse length to allow time for penetration of the fusion front through the workpiece. Figure 12-2 shows depth of melting in stainless steel as a function of time. Good fusion can be achieved over a range of pulse lengths if the laser energy is carefully controlled. For pulses shorter than 1 msec, surface vaporization is difficult to avoid.

We have so far emphasized limitations on the penetration depth that can be attained with laser welding. Now we shall consider how these limitations can be overcome. With high-power continuous lasers, different physical phenomena become important, and deep penetration welding becomes possible. Deep penetration welding has been achieved with multikilowatt CO_2 and Nd:YAG lasers. Samples of steel up to 2 in. thick have been welded. The welding apparently occurs because of "keyholing," in which a hole is opened up so that laser energy can penetrate deep into the workpiece and be absorbed there. Flow of molten material closes the hole after passage of the beam to a new area. This phenomenon is similar to electron beam welding with deep penetration. In this regime of high laser power, one is no longer limited by considerations of heat flow such as were described earlier. Impressive seam welding rates in relatively thick materials have been observed.

As a specific example, ¼-in.-thick stainless steel has been seam welded with 3.6 kW of laser power at a speed of 50 in./min. The weld was of good quality, with no damage outside the fusion zone. Deep penetration laser welding has reached production status in selected cases. A specific example will be described in Chapter 14.

Another important parameter that affects laser welding is the reflectivity of the workpiece surface. It defines the fraction of the incident light that is absorbed and contributes to heating effects. The reflectivity is defined as the ratio of the radiant power reflected from the surface to the radiant power incident on the surface. Thus, the reflectivity is a dimensionless number between zero and unity.

The reflectivity of several metals as a function of wavelength is shown in Figure 12-3. These curves represent typical smooth surfaces. The exact value of the reflectivity is a function of variable conditions, including surface finish and state of oxidation. Thus, the values in Figure 12-3 cannot be interpreted as being exact values for a specific metal.

The figure does show several important features. Copper, for example, has low reflectivity in the blue portion of the visible spectrum and a relatively high value in the red portion. This accounts for the color of copper metal. Metals such as silver and aluminum have reflectivities that are fairly uniformly high throughout the visible spectrum. This accounts for the whitish appearance of these metals. Ferrous metals (steel and nickel alloys) have typically lower reflectivity through the visible spectrum. Thus, they usually appear duller than metals like silver.

The reflectivity of all metals becomes high at long infrared wavelengths. At wavelength longer than 5 μm, the reflectivity strongly depends on electrical conductivity. Metals with high electrical conductivity have the highest values of infrared reflectivity. Thus, the reflectivity of gold is higher than that of aluminum, which in turn is higher than that of steel.

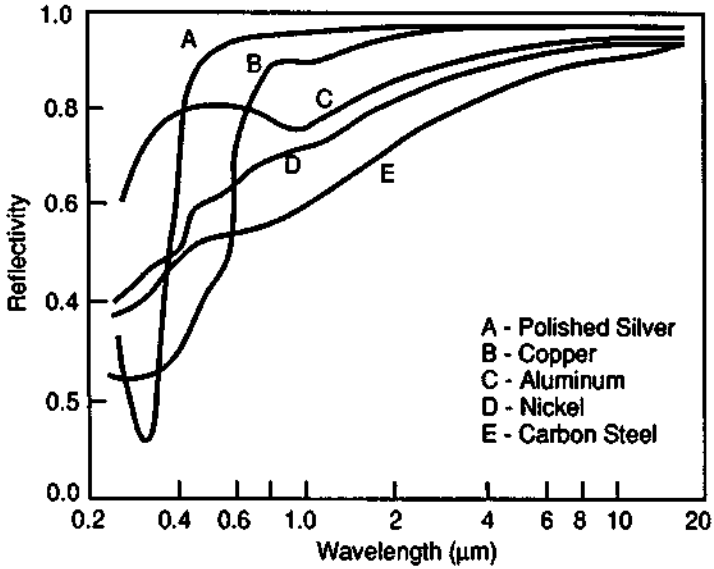


Figure 12-3 Reflectivity as a function of wavelength for several metals.

The amount of light absorbed by a metallic surface is proportional to $1 - R$, where R is the reflectivity. At the CO_2 laser wavelength of 10.6 μm, where R is close to unity, $1 - R$ becomes small. This means that only a small fraction of the light incident on the surface is absorbed and is available for heating the workpiece.

The difference in the value of R becomes important at long wavelengths. For copper or silver at 10.6 μm, $1 - R$ is about 0.02, whereas for steel it is about 0.05. Steel then absorbs about 2.5 times as much of the incident light as silver or copper. In practice, this means that steels are easier to weld with a CO_2 laser than are metals such as aluminum or copper.

The wavelength variation is also important. At shorter wavelengths, the factor $1 - R$ is much higher than at long infrared wavelengths. For example, the factor $1 - R$ for steel is about 0.35 at 1.06 μm, about 7 times as great as its value at 10.6 μm. This means that, at least initially, 7 times as much light is absorbed from a Nd:YAG laser than from a CO_2 laser for equal irradiance from the two lasers. In some cases it will be easier to carry out welding operations with a shorter-wavelength laser because of the increased coupling of light into the workpiece.

But other factors can serve to increase the effective coupling at longer wavelengths. The initially high reflectivity may decrease during the laser interaction. In one experiment [2], an aluminum surface exposed to a Nd:glass laser pulse with an irradiance of 10^7 W/cm^2 had an original reflectivity around 70 percent. When the laser pulse began, the reflectivity dropped steadily to 20 percent after 200 μsec and then remained fairly constant for the remainder of the 1 msec pulse. Therefore, the laser energy was coupled effectively into the material, because the average reflectivity during the pulse was low.

Surface reflectivity is especially important for metalworking with CO_2 lasers. High surface reflectivity at 10.6 μm sometimes leads to difficulties in welding with continuous CO_2 lasers. It is hard to couple the laser energy into the workpiece. Coating of the surface with an absorbing coating is not always effective because of poor thermal coupling between the coating and the workpiece. The coating may be rapidly evaporated, leaving the underlying metal untouched. The reflectivity problem has been a barrier to the application of CO_2 lasers to welding of metals such as gold or copper. Ferrous metals, like steels, have lower reflectivity at 10.6 μm and are better candidates for welding with CO_2 lasers. The decrease of reflectivity that can occur during the laser pulse is especially helpful for welding with CO_2 lasers. Because of this decrease, which occurs when a high-power CO_2 laser beam irradiates a surface, CO_2 lasers do indeed have many practical uses for welding.

Figure 12-4 shows some data on the reflectivity of a stainless steel surface struck by a 200-nsec-duration pulse from a CO_2 TEA laser, which delivered an irradiance of 1.5×10^8 W/cm^2 to the target. The reflectivity dropped rapidly for a few hundred nanoseconds. These date show how the reflectivity can decrease during the time that the surface is irradiated so as to increase the effective absorption of the surface.

To summarize these comments, CO_2 lasers can be effective for welding of ferrous metals, by virtue of the decrease of reflectivity that occurs during irradiation. But for high-conductivity metals, like aluminum and copper, the initial coupling of

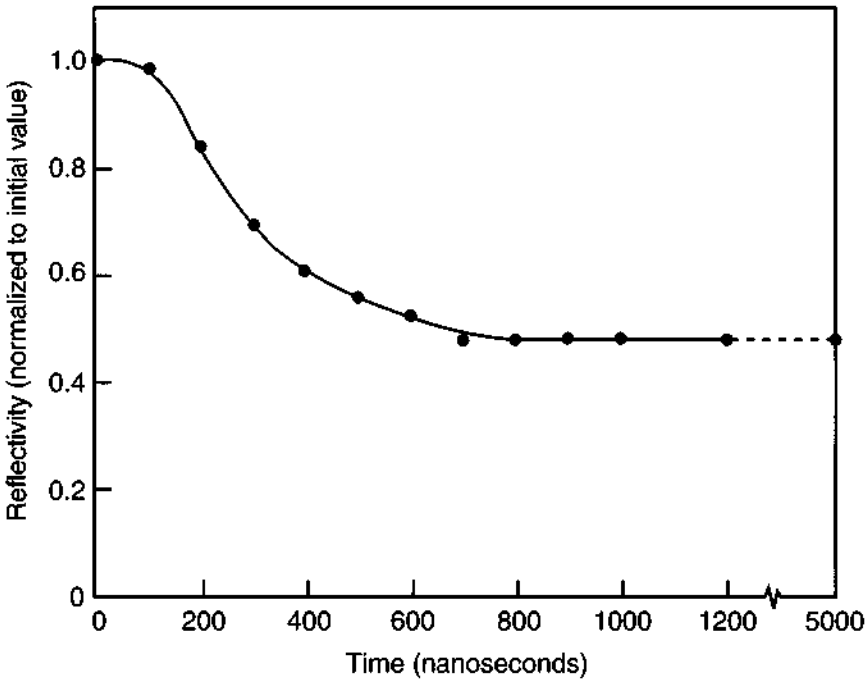


Figure 12-4 Specular reflectivity at 10 μm as a function of time for a stainless steel surface struck by a CO_2 TEA laser pulse delivering 1.5×10^8 W/cm^2 in a pulse 200 nsec long.

CO_2 laser energy into the surface is too low, and the use of a shorter-wavelength laser is indicated.

We have so far considered surface vaporization as undesirable. But in many cases one does want vaporization, as in cutting and hole drilling. When the laser delivers high irradiance to a surface, the metallic surface quickly reaches its vaporization temperature, and a hole is vaporized into the surface.

The lasers most useful for hole drilling have pulse durations in the 100 μsec to 10 msec regime, a time scale that allows the surface to heat to the vaporization temperature and remain there for some time. When a normal pulse laser beam (duration around 1 msec) interacts with a surface, the process of material removal involves conventional melting and vaporization. The time scale of the interaction is long enough that the vaporized material can flow away from the interaction region. The material is removed without further interaction with the laser beam. Vaporization occurs at a continually retreating surface, with the laser energy delivered to the bottom of the developing hole. Under typical conditions, the vaporization temperature is reached within a very short time. The time t_B to reach the vaporization temperature is given by the following equation:

$$t_B = \frac{\pi}{4} \frac{K\rho c}{F^2} (T_B - T_0)^2 \tag{12.2}$$

where K, ρ, and c respectively are the thermal conductivity, density, and heat capacity of the material, T_B is its vaporization temperature, T_0 is the ambient temperature, and F is the absorbed irradiance.

Table 12-2 shows values calculated using Equation (12.2) for absorbed laser irradiance of 10^5–10^7 W/cm^2. These are typical values for the irradiance at a target surface using common lasers and simple focusing optics. Because of surface reflection, the absorbed irradiance may be lower than the incident irradiance. The values of time in Table 12-2 may be very short. At high values of absorbed irradiance, the surface may begin to vaporize very quickly.

The values in Table 12-2 were calculated using the assumption that uniform laser irradiance is absorbed at an infinite plane surface. Figure 12-5 shows the effect of a finite laser spot size. The figure shows the absorbed laser irradiance required for surface vaporization to begin as a function of laser pulse duration and the radius of the beam on an aluminum surface. For short pulse durations, there is insufficient time for transverse thermal conduction in the plane of the surface, and the result is independent of spot size. It also corresponds to the results of Equation (12.2). For longer pulse durations, transverse thermal conduction becomes important, and the results depend on spot size. For small spot sizes, the transverse thermal gradients are higher, and heat is conducted out of the focal area more quickly. Thus, a small focal spot requires higher irradiance to produce vaporization.

Before vaporization begins, the surface must first start to melt. Because of the great speed with which vaporization begins, there is not time for much material to melt. Thus, at high laser irradiance (>10^6 W/cm^2), the dominant physical process is vaporization, and the role of melting tends to be less significant. After the surface reaches its vaporization temperature, the laser continues to deliver more energy. This energy supplies the latent heat of vaporization. Material is then removed from the target as vapor. The result is the drilling of a hole into the target.

Table 12-2 **Time to Reach Vaporization Temperature**

	Absorbed Laser Irradiance (W/cm^2)		
Metal	**10^5**	**10^6**	**10^7**
Lead	118 μsec	1.18 μsec	12 nsec
Zinc	128 μsec	1.28 μsec	13 nsec
Magnesium	245 μsec	2.45 μsec	24.5 nsec
Titanium	319 μsec	3.19 μsec	31.9 nsec
Chromium	1.54 msec	15.4 μsec	154 nsec
Nickel	1.84 msec	18.4 μsec	184 nsec
Iron	1.86 msec	18.6 μsec	186 nsec
Aluminum	3.67 msec	36.7 μsec	367 nsec
Molybdenum	5.56 msec	55.6 μsec	556 nsec
Copper	8.26 msec	82.6 μsec	826 nsec
Tungsten	10.46 msec	104.6 μsec	1046 nsec

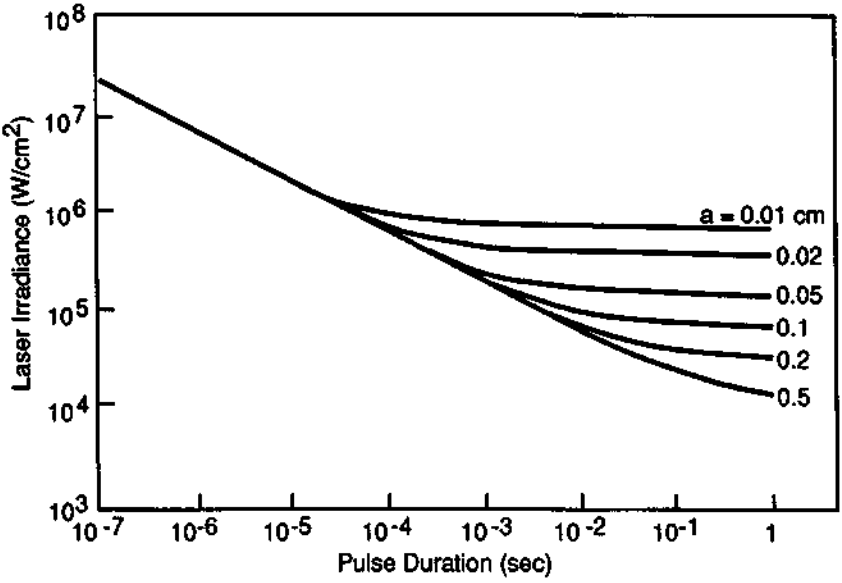


Figure 12-5 Laser irradiance required to raise a massive aluminum surface to its vaporization temperature as a function of the pulse duration and the Gaussian radius of the laser beam.

The time to reach the vaporization temperature is often a small fraction of the laser pulse duration. In that case, an equilibrium condition may be reached, so that the vaporizing surface retreats at a steady rate v_B equal to

$$v_B = \frac{F}{\rho[L + c(T_B - T_0)]} \tag{12.3}$$

where L is the specific latent heat of vaporization. For reasonable conditions, this velocity may persist throughout most of the duration of a normal laser pulse. For a typical metal and a 1-msec-duration laser pulse, one derives a hole depth around 1 mm. This is a typical depth attained with small normal pulse solid state lasers, like ruby.

It is important to emphasize here that the depths of holes drilled by lasers are limited. Typically, an ordinary small laser will drill holes only one or a few millimeters deep.

Additional results for depth vaporized under various conditions will be presented in Chapter 16. We will mention here that a very important factor is the latent heat of vaporization. Metals with low heat of vaporization are vaporized to a larger extent than metals with high heat of vaporization, like tungsten. Under many conditions, the heat of vaporization of the workpiece is the dominant parameter in determining how much material is removed. In fact, for high values of laser irradiance, only a small amount of the incident energy is lost via thermal conduction out of the irradi-

ated spot. Under these conditions, the depth D of the hole may be approximated by the simple equation

$$D = \frac{Ft_p}{L\rho} \tag{12.4}$$

where t_p is the pulse duration and L is the latent heat.

The amount of material vaporized depends of the exact conditions of the laser irradiation. Thus, results quoted by different workers may differ.

During laser-based hole drilling, not all of the ejected material is vaporized. When a hole is produced in the target, the vapor builds up a pressure that causes a flow toward the exit aperture of the crater. The flow can carry along some of the molten material at the edges of the crater. This flushing process removes some of the material as unvaporized droplets of molten material. These ejected hot molten droplets appear like a shower of sparks. This is a common feature associated with laser-based hole drilling. Because one does not have to supply the relatively large amount of energy for the latent heat of vaporization, this flushing process results in a mass removal that is larger than if only vaporization occurred.

A scanning electron microscope photograph of an area impacted by several successive CO_2 laser pulses of 100 nsec duration is shown in Figure 12-6. It is apparent that even for this short pulse duration, molten material flowed within the crater and splattered around the crater edges.

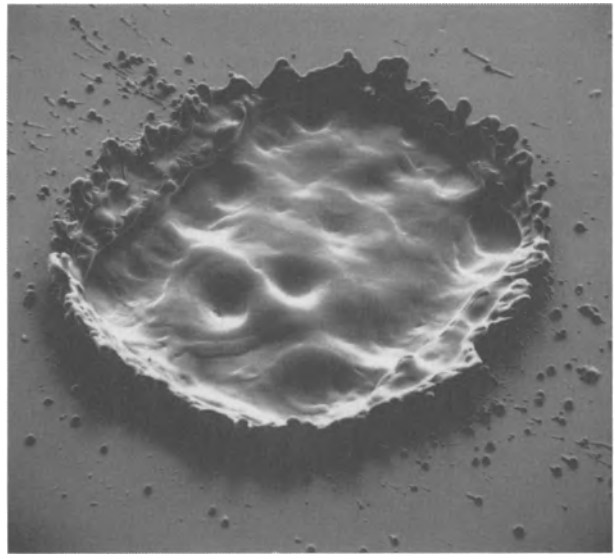


Figure 12-6 Scanning electron microscope photograph of a crater produced by the impact of several successive TEA laser pulses delivering 3×10^8 W/cm^2 to a stainless steel surface. The width of the area shown is approximately 1.8 mm.

Under some conditions, most of the material may be removed as liquid. Figure 12-7 shows data on the fraction of the material ejected as liquid by a Nd:glass laser pulse with 30 kW power. Early in the pulse, most of the material was removed as vapor, but after a few hundred microseconds, about 90 percent of the material removal occurred as liquid droplets.

One might think that one should use lasers with very high peak power in order to increase material removal. Paradoxically, it is not the highest laser powers that are optimal for material removal. The very high powers from a Q-switched laser vaporize a small amount of material and heat it to a high temperature. Early in the laser pulse, some material is vaporized from the surface. The vaporized material is slightly thermally ionized and absorbs some of the incident light. This heats the vapor more, producing more ionization and more absorption in a feedback process. Our earlier assumption that the laser light does not interact with the vaporized material is no longer valid if the irradiance becomes very high. Rather, the vaporized material does interact and absorb the incoming laser beam, so that the surface is shielded from the laser light.

Thus, new physical processes become important as the irradiance becomes very high. This is shown schematically in Figure 12-8, which shows depth vaporized as a

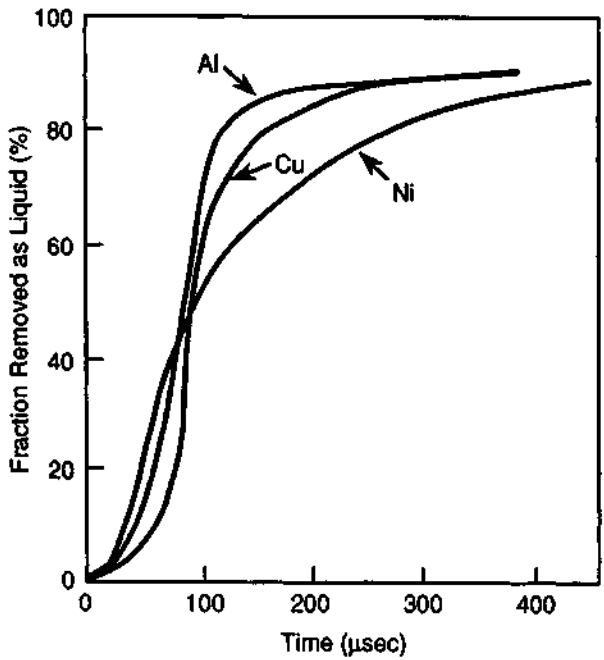


Figure 12-7 Fraction of material removed in liquid form. Results are shown as a function of time for several metals struck by a Nd:glass laser pulse. (From M. K. Chun and K. Rose, *J. Appl. Phys.* **41**, 614 (1970).)

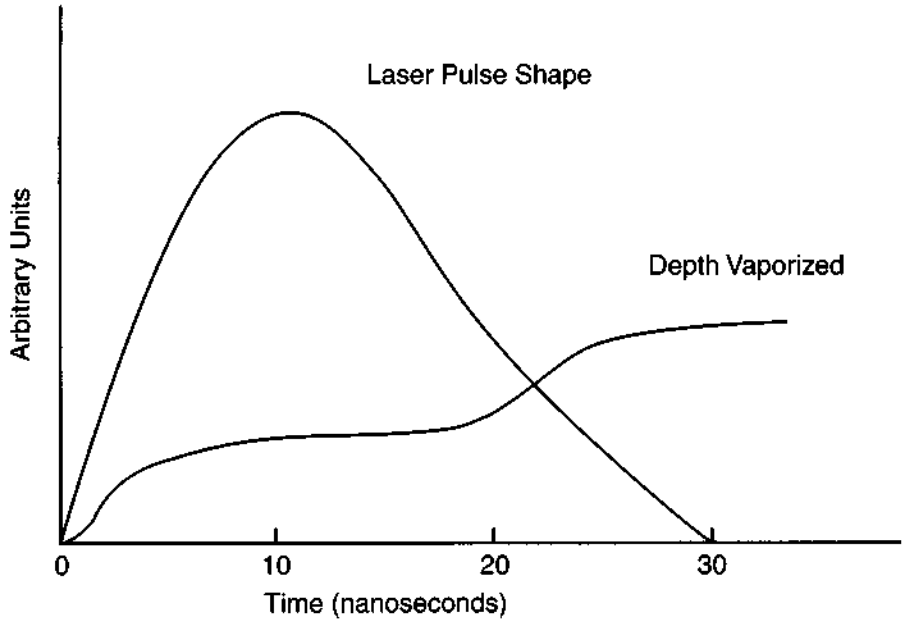


Figure 12-8 Schematic representation of the depth vaporized in a metallic target as a function of time by a 30-nsec-duration laser pulse with the indicated temporal profile. The effect of shielding of the target surface by blowoff material produced early in the laser pulse is apparent. (From J. F. Ready, *Effects of High-Power Laser Radiation*, Academic Press, New York, 1971.)

function of time. The laser pulse shape is also shown for comparison. This is a typical pulse shape for Q-switched lasers. Early in the pulse, the surface starts to vaporize. Then the vaporized material, heated and ionized by the laser, forms a hot, opaque, ionized plasma, which absorbs essentially all of the incoming laser light.

The flat portion of the curve represents the period when the surface is shielded by the plasma, so that vaporization ceases. Finally, late in the pulse, the plasma has expanded and become transparent again. Light can again reach the surface, and some additional material is vaporized. Because of these effects, the amounts of material that can be removed by short-duration, high-power pulses, as from a Q-switched laser, are limited. Such lasers are not well suited for hole drilling or for cutting.

The shielding of the target by a hot opaque plasma leads to a phenomenon called a laser-supported absorption (LSA) wave. The LSA wave is a plasma that is generated above the target surface and propagates backward along the beam path toward the laser. It is accompanied by a loud noise and a bright flash of light. Thus, the LSA wave makes an impressive demonstration, but while present it effectively shields the target surface and reduces the material removal. It can also drive a shock wave into the target.

Some experiments [3] directly demonstrating the LSA wave were performed with a CO_2 gas dynamic laser that delivered 5-msec-duration pulses with an irradiance of 1–2 MW/cm^2. When aluminum 2024 alloy was irradiated, the beam coupled to the target surface and generated a plume of incandescent particles that extended above the surface toward the laser. The form of the plume, as photographed with a high-speed framing camera, is shown in Figure 12-9. Approximately 1.71 msec after the start of the pulse, the plume had started to decouple from the surface because of the attenuation of the beam by the plume. About 1 msec later, the plume was completely decoupled from the target At that time, very little energy was reaching the target surface.

When this occurred, the target no longer supplied material to the plasma, so that the plasma began to dissipate and its absorption decreased. In addition, the plume had propagated back toward the laser to a position where the laser irradiance was lower, and the laser could no longer sustain the plasma. At 4.54 msec, the laser beam again coupled to the surface and remained coupled until the pulse ended at 5.14 msec. The plume then quickly dissipated until another laser pulse struck the surface.

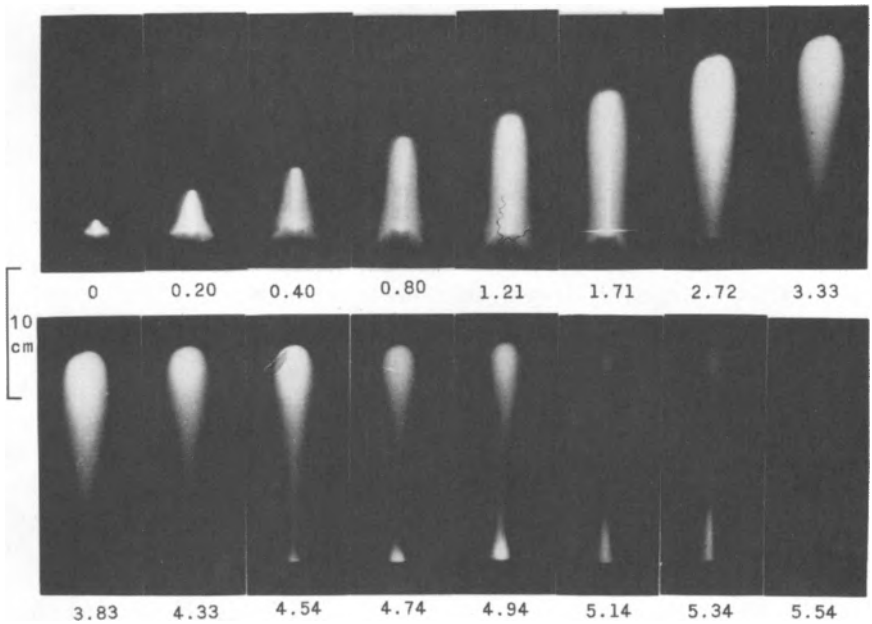


Figure 12-9 Vaporized material emitted from an aluminum surface struck by a CO_2 laser pulse delivering 1.5×10^6 W/cm^2 The beam was incident from the top. Selected frames of high-speed photography of the plume show decoupling of the laser radiation from the surface after 2.72 msec. Each frame is identified by the time in milliseconds after the start of the laser pulse. (From R. L. Stegman, J. T. Schriempf, and L. R. Hettche, *J. Appl. Phys.* **44**, 3675 (1973). Photograph courtesy of J. T. Schriempf.)

Different target materials gave different responses. An MgO target gave approximately four cycles of plume generation, decoupling, plasma dissipation, and recoupling during the 5 msec pulse. A pyroceram target went through about eight cycles. But graphite and fiberglass–epoxy targets remained continuously coupled throughout the pulse. Thus, the optimum pulse duration for a laser in a material removal operation depends on the target material. If the pulse duration exceeds the optimum value for the particular target, laser energy is wasted in the LSA wave.

The effect of the shielding by the LSA wave is demonstrated in Figure 12-10, which shows relative amounts of material removal from an alumina target struck by a CO_2 laser pulse with 1 μsec duration. This figure shows specific mass removal, that is, mass removal per unit laser energy, which is a measure of the efficiency of the material removal process. As irradiance increases, the specific mass removal increases, up to about 10^7 W/cm^2. Above 10^7 W/cm^2, the laser-supported absorption wave develops and shields the target so that the specific mass removal is reduced. The specific mass removal then decreases with further increases in irradiance.

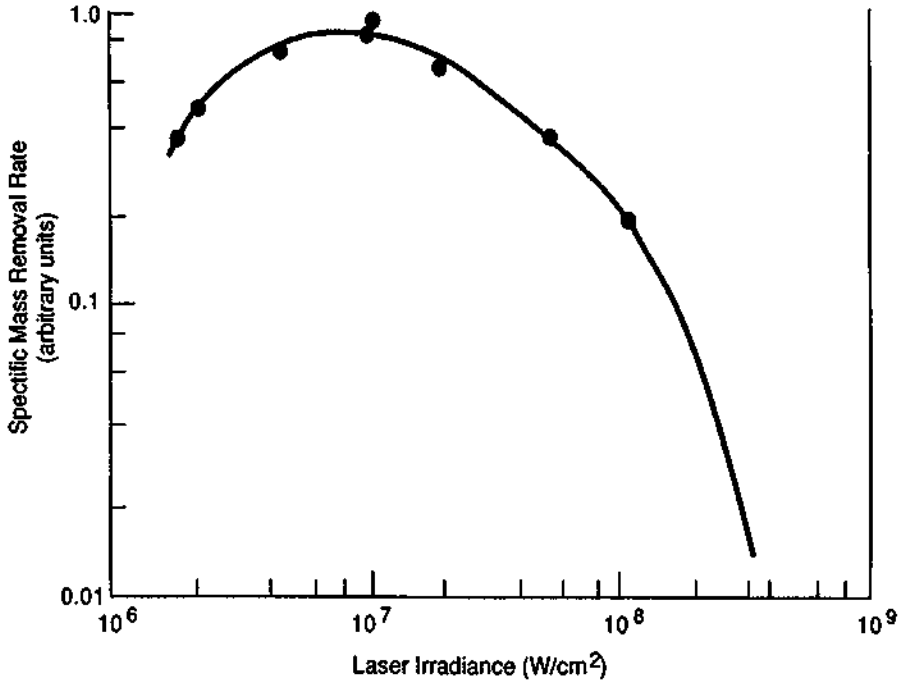


Figure 12-10 Specific mass removal from an alumina surface struck by a CO_2 laser pulse with 1 μsec duration as a function of laser irradiance. The specific mass removal is the amount of material removed per unit incident energy. It gives a measure of the efficiency of the utilization of incident energy for material removal.

Absorption of laser energy can also produce large pressure pulses in the target. One mechanism by which pressure pulses are produced involves evaporation of material from the surface, with recoil of the surface in response to the momentum of the vaporized material. Larger pressure pulses are developed when the LSA wave is kindled. The hot plasma, as it expands, can drive a shock wave into the target. Peak shock pressures of tens of kilobars have been measured in metallic materials irradiated by high-power Q-switched Nd:glass lasers. The pressure pulse can be large enough to cause spallation of the back surface of a thin metallic plate.

The expanding plasma also imparts an impulsive motion to the target as a whole. Considerable experimental work has been devoted to measuring the impulse transmitted to the target. This work has mostly been done at high values of laser irradiance, where a large amount of material is removed from the surface.

Variation of momentum transfer to the target as a function of laser irradiance has been extensively studied. For example, the momentum delivered by a focused ruby laser pulse to a target in vacuum was measured by suspending the target by a thread to form a simple pendulum. The motion of the pendulum was determined with a calibrated microscope. The specific momentum transfer as a function of laser irradiance for several materials is shown in Figure 12-11. The specific momentum transfer (also called specific impulse) is the momentum transferred per unit incident energy. It has units of dyn-sec/J. The observed values in the figure are much higher than what would be transferred through photon reflection. The figure shows that there is an optimum irradiance for each material that gives the maximum specific impulse. At values below the optimum, some energy is lost by thermal conduction. Above the optimum, some energy goes into increasing the temperature and ionization state of the blowoff material. This is less effective for transferring momentum than if the same amount of energy were used to vaporize a larger number of atoms.

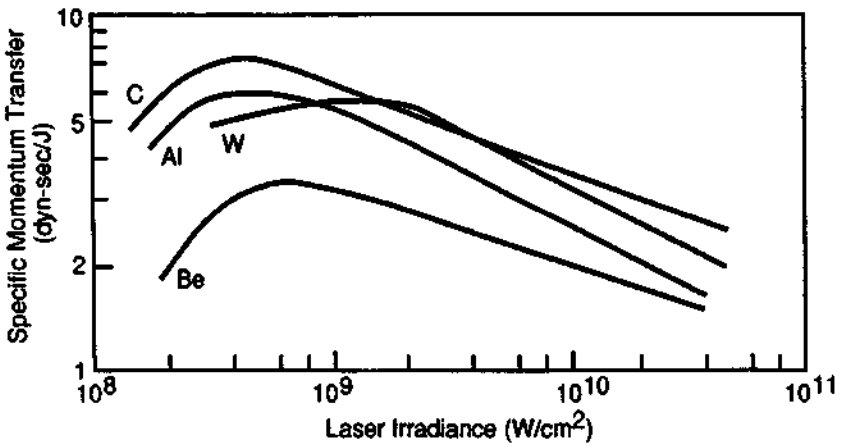


Figure 12-11 Specific momentum transfer to several materials from a Q-switched ruby laser pulse. (From D. W. Gregg and S. J. Thomas, *J. Appl. Phys.* **37**, 2787 (1966).)

At very high values of irradiance ($\geq 10^{13}$ W/cm^2), additional physical phenomena, dominated by plasma effects, become important. The solid surface disappears rapidly, and a plasma (ionized gas) is formed. The physical effects involve things such as collective oscillations of the charged particles in the plasma. This is a regime that is of high interest to people working in the field of laser-assisted thermonuclear fusion research, where very large lasers deliver irradiance in excess of 10^{16} W/cm^2 to targets. But it is a regime that has not been of importance for industrial applications.

We have now described the important physical phenomena that occur during the interaction of high-power laser radiation with surfaces, and we are in a position to summarize. The physical phenomena are sketched if Figure 12-12. The top left portion of the figure indicates absorption of the incident laser light according to the exponential absorption law

$$I(x) = I_0 e^{-\alpha x} \tag{12.5}$$

where $I(x)$ is the light intensity at depth x, I_0 is the incident light intensity, and α is the absorption coefficient. For purposes of this figure, we neglect the fraction of the light that is reflected. For metals, the absorption coefficient is of the order of 10^5 cm^{-1}. Thus, the energy is deposited in a layer about 10^{-5} cm thick. The light energy is transformed into heat energy essentially instantaneously, in a time less than 10^{-13} sec. Thus the laser energy may be regarded as an instantaneous surface source of heat.

The heat energy then penetrates into the target by thermal conduction. When the surface reaches the melting temperature, a liquid interface propagates into the material, as indicated in the top right portion of Figure 12-12. With continued irradiation,

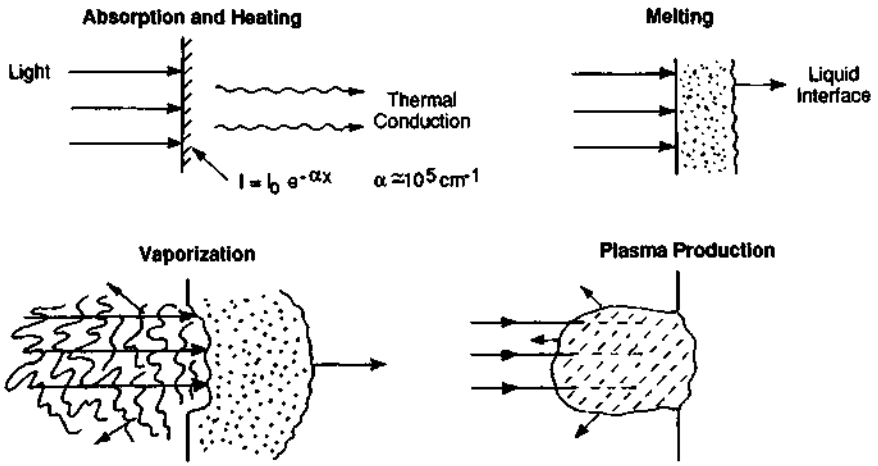


Figure 12-12 Physical phenomena occurring when a high-power laser beam strikes an absorbing surface.

the material begins to vaporize, as indicated in the bottom left portion of Figure 12-12, and a hole begins to be drilled. If the irradiance is high enough, absorption in the blowoff material leads to a hot opaque plasma. The plasma can grow back toward the laser as an LSA wave. The plasma absorbs the light and shields the surface, as shown in the bottom right portion of Figure 12-12.

The ranges of laser irradiance for which individual processes dominate the interaction are given in Table 12-3. Values are stated for two wavelength regions: the visible and near infrared region (around 0.5–1 μm) and the far infrared region near 10 μm. The values in the table are approximate and will vary according to the exact parameters of the laser irradiation, such as pulse duration, target properties, and the like. At relatively low irradiance, melting is the main effect. At somewhat larger irradiance, vaporization becomes the most important effect. This is conventional vaporization, with minimal interaction between the incident light and the vaporized material.

At still higher irradiance, LSA waves are kindled and dominate the physical processes, whereas vaporization is diminished. The threshold for kindling the LSA waves are those appropriate to one specific case, namely, a titanium target with a laser pulse duration in the microsecond regime. The threshold will vary as the circumstances change. But the numbers in Table 12-3 will serve to identify an order of magnitude at which certain types of interaction occur. The LSA wave dominates at a lower value of irradiance for far infrared lasers than for visible and near infrared lasers.

Finally, when the irradiance becomes very high, additional absorption mechanisms may become operative, like excitation of collective oscillations of the electrons and ions in the plasma. Investigation of these effects has been performed by workers in the laser-assisted thermonuclear fusion community, but these interactions at very high irradiance have not found industrial applications.

We note that the regime of greatest interest for material processing applications (cutting, drilling, etc.) lies below the threshold for kindling of the LSA waves. Below this threshold, the laser energy is used mainly for changing (melting or vaporizing) the target, whereas above the threshold, the energy goes mainly into the LSA wave or into plasma effects. The ignition of the LSA wave may appear spectacular, with a loud noise and a bright flash of light, but the solid target surface is relatively little affected. Thus, for most of the applications that we will describe in the follow-

Table 12-3 **Approximate Ranges of Laser Irradiance at which Various Processes Dominate the Laser–Surface Interaction**

Process	Range for visible and near infrared laser (W/cm^2)	Range for far infrared laser (W/cm^2)
Melting	$\sim 10^5$	$\sim 10^5$
Vaporization	10^6–1.5×10^8	10^6–2.5×10^7
Laser-supported absorption (LSA) wave	$>1.5 \times 10^8$	$>2.5 \times 10^7$
Plasma-collective effects	$\geq 10^{13}$	$>10^{13}$

ing chapters, the production engineer will want to work at values of laser irradiance below the threshold for LSA waves.

One possible exception, where the threshold for LSA wave ignition is exceeded deliberately, involves shock hardening. Certain alloys may be hardened by passage of a shock wave through the material. There have been some studies of hardening aluminum alloys via the shock wave that accompanies the LSA wave. Other studies have indicated that laser shock processing, as it is called, can increase fatigue resistance of aluminum alloys.

Laser shock processing has been considered for applications in which shot peening has been used to protect areas for which fatigue is critical. Laser shock processing has not yet reached production status, but it remains under investigation [4].

We summarize these phenomena in Figure 12-13, which identifies various regimes of interaction and their potential applications. The figure defines these regimes in terms of irradiance and the duration of the interaction. The ordinate represents the pulse duration for a pulsed laser (or the time that the beam dwells on a spot for a continuous laser). Below the line marked "No Melting," the surface is not heated to the melting point. In this region, one may have heat treating applications. In the region marked "Welding," one obtains a reasonable depth of molten material, and welding applications are possible. Above the line marked "Surface Vaporization," the surface begins to vaporize and welding applications are less desirable. To the left of the welding region, the penetration of the fusion front is small because of the short interaction time. To the right of the welding region, the heat spreads over a

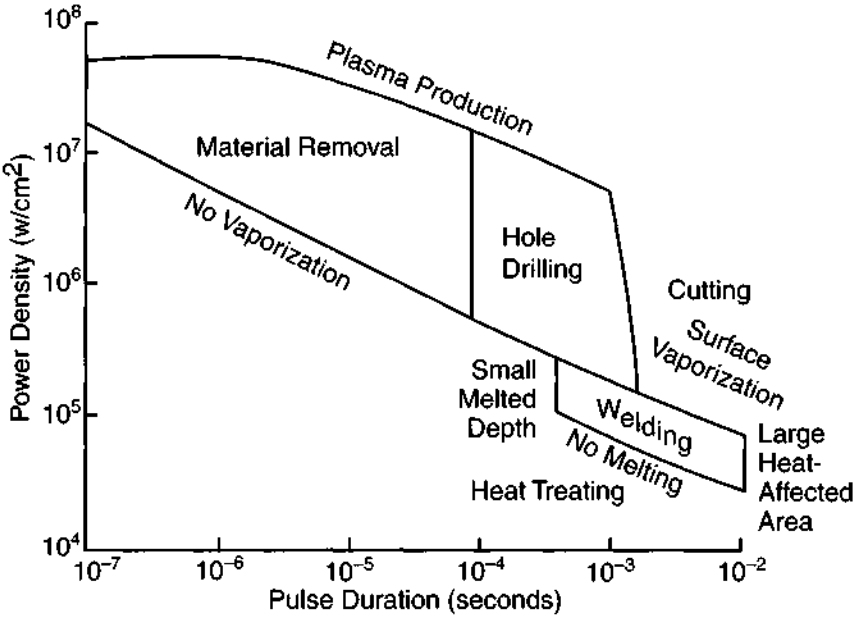


Figure 12-13 Regimes of laser irradiance and interaction time for material processing applications.

broad area, and the desirable feature of localized heating is lost. Thus, welding operations usually require careful control to remain within this process window. Similarly, the figure identifies regimes useful for cutting, hole drilling, and material removal for small amounts of material, such as vaporization of thin films, that is, trimming. Above the line marked "Plasma Production," the LSA wave develops. The only potential application identified in this region has been shock hardening.

The regions identified in Figure 12-13 are not exact; they will vary with target material, laser wavelength, and so forth. Still, they define regimes of laser parameters where certain applications are most likely to be productive. The engineer desiring to apply a laser in a specific material processing application must identify the process parameters suitable for that particular application.

In the following chapters, we will discuss the applications of lasers for the various types of material processing applications in more detail.

References

[1] M. I. Cohen, *J. Franklin Inst.* **283**, 271 (1967).
[2] M. K. Chun and K. Rose, *J. Appl. Phys.* **41**, 614 (1970).
[3] J. T. Schriempf, R. L. Stegman, and G. E. Nash, *IEEE J. Quantum Electron.* **QE-9**, 648 (1973).
[4] A. H. Clauer, *Industrial Laser Review*, p. 7 (March 1996).

Selected Additional References

F. J. Allen, Surface Temperature and Deposition of Beam Energy for a Laser-Heated Target, *J. Appl. Phys.* **42**, 3145 (1971).
J. A. Fox and D. N. Barr, Laser-Induced Shock Effects in Plexiglas and 6061-T6 Aluminum, *Appl. Phys. Lett.* **22**, 594 (1973).
J. N. Gonsalves and W. W. Duley, Interaction of CO_2 Laser Radiation with Solids. I. Drilling of Thin Metallic Sheets, *Can. J. Phys.* **49**, 1708 (1971).
J. E. Lowder and J. C. Pettingill, Measurement of CO_2-Laser-Generated Impulse and Pressure, *Appl. Phys. Lett.* **24**, 204 (1972).
S. A. Metz *et al.*, Effect of Beam Intensity on Target Response to High-Intensity Pulsed CO_2 Laser Radiation, *J. Appl. Phys.* **46**, 1634 (1975).
R. Miller and T. DebRoy, Energy Absorption by Metal-Vapor-Dominated Plasma during Carbon Dioxide Laser Welding of Steels, *J. Appl. Phys.* **68**, 2045 (1990).
C. R. Phipps, Jr., *et al.*, Impulse Coupling to Targets in Vacuum by KrF, HF, and CO_2 Single-Pulse Lasers, *J. Appl. Phys.* **64**,1083 (1988).
J. F. Ready, Effects due to Absorption of Laser Radiation, *J. Appl. Phys.* **36**, 462 (1965).
J. F. Ready, *Effects of High-Power Laser Radiation*, Academic Press, New York, 1971, Chapter 3.
J.F. Ready, Impulse Produced by the Interaction of CO_2 Laser Pulses, *Appl. Phys. Lett.* **25**, 558 (1974).
J. F. Ready, Change of Reflectivity of Metallic Surfaces during Irradiation by CO_2-TEA Laser Pulses, *IEEE J. Quantum Electron.* **QE-12**, 137 (1976).
J. F. Ready, Material Processing—An Overview, *Proc. IEEE* **70**, 533 (1982).
E. N. Sobol, *Phase Transformations and Ablation in Laser-Treated Solids*, Wiley, New York, 1995.
M. Sparks, Theory of Laser Heating of Solids: Metals, *J. Appl. Phys.* **47**, 837 (1976).
M. von Allmen, Laser Drilling Velocity in Metals, *J. Appl. Phys.* **47**, 5460 (1976).
R. E. Wagner, Laser Drilling Mechanics, *J. Appl. Phys.* **45**, 4631 (1974).
T. E. Zavecz, M. A. Saifi, and M. Notis, Metal Reflectivity under High-Intensity Optical Radiation, *Appl. Phys. Lett.* **26**, 165 (1975).

Chapter 13 | Laser Applications in Material Processing

Material processing refers to a variety of industrial operations in which the laser operates on a workpiece to modify it, for example, by melting it or removing material from it. Some of the possible applications include welding, hole drilling, cutting, trimming of electronic components, heat treating, and alloying.

Properties of laser light that enable material processing applications are its collimation, radiance, and focusability. Because of these properties, laser light can be concentrated by a lens to achieve extremely high irradiance at the surface of a workpiece.

Let us compare a CO_2 laser and a large mercury arc lamp, both operating at an output of 1000 W. If the CO_2 laser beam has a divergence angle of 0.01 rad, it can be focused by a lens with a focal length of 1 cm to a spot 0.01 cm in diameter. This diameter D is calculated using the equation

$$D = f\theta \tag{13.1}$$

where f is the focal length of the lens and θ is the beam divergence angle. All the laser light can be easily collected and delivered to this small spot. The irradiance I in the focal area is approximately

$$I \approx \frac{1000\ \text{W}}{(0.01)^2\ \text{cm}^2} = 10^7\ \text{W/cm}^2 \tag{13.2}$$

(This approximation neglects a factor of $\pi/4$, which is of the order of magnitude of unity.) Thus, a fairly ordinary laser can easily produce extremely high irradiance.

The mercury arc lamp, with the same output, is a large lamp that requires a large power supply. It emits power into a solid angle of 4π steradians (sr). Because not all the light can be collected by the focusing optics, and because the large divergence of the light leads to a much larger focal area, the irradiance from the lamp is much lower. It can be shown (although we will not go through the details here) that the same lens, with a 1 cm focal length, will deliver an irradiance of 100 W/cm^2 to a surface. Thus, the irradiance produced by the laser is 100,000 times as large, although

the initial power was the same. This illustration shows why laser radiation is useful for material processing.

Pulsed lasers can of course deliver much higher values of irradiance in a short pulse. For example, small easily available Nd:glass or ruby lasers can deliver 1-msec-duration pulses to a surface with irradiance exceeding 10^9 W/cm^2. No other energy source (except a large electron beam machine) can produce such high values of irradiance.

The essential feature of laser material processing is very localized delivery of high irradiance. Thus, lasers can be used for precision processing. They produce very large heating rates in the affected volume, while areas not far away are not heated. This leads to small heat-affected zones. It makes laser processing advantageous when one desires to affect only a localized region.

In the early history of lasers, the operating costs were relatively high, and lasers did not compete well economically with other processing techniques. They did find many applications where their localized heating and large irradiance offered technical advantages. Thus, they performed specialized processing applications, for which conventional tools suffered some problem.

The continued development of lasers has changed this situation. Today, lasers still offer technical advantages in many cases, and they solve many difficult problems in material processing. But they can also compete economically in many situations with other techniques. They offer easy fixturing in many cases and high throughput, and they can displace other established techniques simply because laser processing is less expensive.

The operating costs for both CO_2 and Nd:YAG lasers may be in the range from five to ten dollars per hour (1996 dollars) for lasers operating at power levels up to a few kilowatts. This includes the costs of electrical power and consumable items, like water filters and arc lamps, but not the costs associated with the workstation, like shielding gas. Figure 13-1 shows an estimate of how operating costs increase with laser output, for CO_2 and Nd:YAG lasers. At a given level of power, the operating cost of the CO_2 laser is somewhat lower than that of the Nd:YAG laser, largely because of the higher efficiency of the CO_2 laser. But because of its shorter wavelength, the Nd:YAG laser beam can be focused to a smaller size than the CO_2 laser, producing a higher irradiance. Thus, some applications may be carried out with lower power from a Nd:YAG laser than from a CO_2 laser. In addition, the use of a Nd:YAG laser offers advantages of smaller size and less expensive optics.

In a laser material processing operation, the lens must be protected from material vaporized from the workpiece. Debris splattered back on the lens can quickly damage the lens, requiring costly replacement. The replacement is especially costly for the zinc selenide lenses often used with CO_2 lasers.

Lenses may be protected by a gas jet that blows the debris away from the lens. In a system using a Nd:YAG laser, an inexpensive replaceable silica cover glass may be used to protect the lens.

Compared with conventional processing technologies, laser processing offers a number of advantageous features:

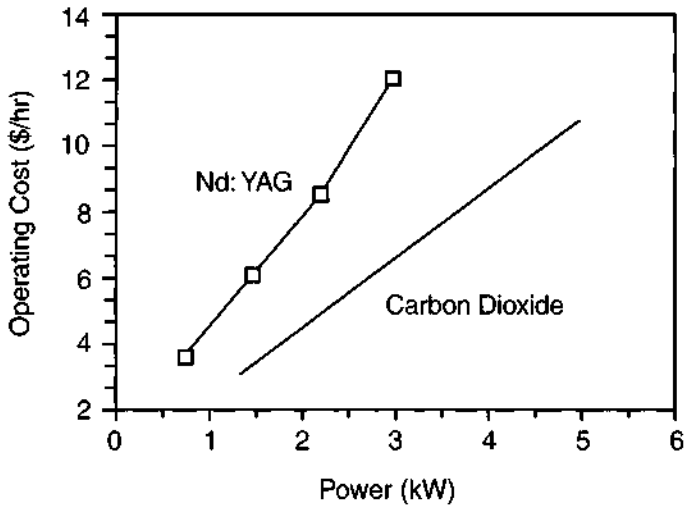


Figure 13-1 Estimated operating cost as a function of laser output power for CO_2 and Nd:YAG lasers. (Taken from information in D. B. Veverko, *Industrial Laser Review*, p. 19 (January 1996).

- Very high values of irradiance and very localized heating can be achieved. A laser can deliver higher irradiance than any other thermal source.
- Heat-affected zones can be very small.
- In many cases, fixturing is easy and fast.
- The processing is easily compatible with automation.
- Setup can be speedy, leading to the capability of very rapid prototyping.
- No vacuum is required. This leads to a capability for rapid throughput. For some applications, a gas jet may be desirable.
- There is no contact of any material with the workpiece, so that contamination problems are reduced.
- Laser processing can work well with some "difficult" materials, such as refractory materials or hard, brittle materials.
- Small hole diameters can be achieved.
- Extremely small welds may be accomplished on delicate objects.
- Inaccessible areas or even encapsulated materials can be reached with the laser beam.

There are some limitations to laser processing:

- The initial capital cost is usually high.
- The depth of penetration in laser welding is limited, except for multikilowatt lasers.

- In laser welding, careful control of the process is required to avoid surface vaporization.
- The depth of penetration for laser-drilled holes is limited, although repeated pulses can increase the depth.
- The quality of laser-drilled holes is less than perfect. The walls of the holes are generally rough. Their cross sections are not completely round, and they taper somewhat from entrance to exit. There is often some recondensation of vaporized material on the walls of the holes and at the entrance to the holes. The control of size and tolerances is not perfect.

Often, the advantages outweigh the limitations. In fact, laser processing has often been used to solve problems that were difficult for conventional processing. Some examples of considerations that might suggest use of laser processing include

- The necessity of producing a small heat-affected zone
- The necessity of avoiding deposition of electrical charge in the workpiece
- The use of difficult materials, for example, hole drilling in ceramics

Laser processing has come to be widely accepted for many industrial applications. It will not replace all the multitudinous welding, cutting, drilling, and trimming applications of industry, but it has established a firm foundation of applications. It has been accepted for applications in which it solves difficult processing problems, and for applications in which it offers high throughput, economical processing, and rapid turnaround.

Of the many different types of lasers that have been developed, only a relatively small number are useful for material processing. The leading contenders are listed in Table 13-1. The CO_2 and Nd:YAG lasers, in their various modes of operation, have dominated laser material processing. Most established material processing applications use one of these two types. Ruby and Nd:glass lasers have long been used

Table 13-1 **Lasers for Material Processing**

Laser	Wavelength (μm)	Operating regimes	Typical applications
CO_2	10.6	Continuous, pulsed, TEA	Welding, drilling, heat treating
Nd:YAG	1.06, 0.532	Continuous, pulsed	Welding, drilling, trimming, marking
Ruby	0.6943	Pulsed	Spot welding, drilling
Nd:glass	1.06	Pulsed	Spot welding, drilling
Alexandrite	0.72–0.78	Pulsed	Drilling
Copper vapor	0.511, 0.578	Pulsed	Drilling
Excimer	0.249	Pulsed	Micromachining, ablation, semiconductor processing
Argon	0.488, 0.5145	Continuous	Semiconductor processing

in applications where a small number of pulses of relatively large energy are required. Applications of the other lasers in the table are developing. The availability of the shorter wavelengths of the excimer lasers offers advantages when fine focusing is needed.

For the two leading types, CO_2 and Nd:YAG, Tables 13-2 and 13-3 give additional information about the power levels used in typical applications, for metals and nonmetals, respectively. The tables present an applications matrix, indicating which types of lasers are commonly used in various applications.

We emphasize that a laser material processing system contains many components other than the laser. The laser is only one element and by itself could not successfully carry out the application. In general, a complete system may have the laser and laser power supply, a gas management system if necessary, probably a beam delivery system, a shutter, beam focusing optics, possibly gas jets, usually a power monitoring device, fixturing for holding and moving the workpiece, sometimes equipment for moving the beam, appropriate safety devices and interlocks, possibly equipment for monitoring the effect of the beam on the workpiece, and often programmable drives for automatically positioning the workpiece and programmable controllers for driving the laser. For the application to be successful, all parts of the system must be carefully designed and chosen.

Table 13-2 **Lasers for Processing Metals**

Laser	Heat treating	Welding	Drilling	Cutting	Marking
Medium-power CO_2 (200–800 W)	—	Yes	Yes	Yes	—
High-power CO_2 (>900W)	Yes	Yes	Yes	Yes	—
Low/medium-power Nd:YAG (<100 W)	—	Yes	Yes	—	Yes
High-power Nd:YAG (>100W)	—	Yes	Yes	Yes	—

Table 13-3 **Lasers for Processing Nonmetals**

Laser	Drilling	Cutting	Scribing	Marking
Low-power CO_2 (20–200 W)	Yes	Yes	Yes	Yes
Medium-power CO_2 (200-800 W)	Yes	Yes	Yes	—
High-power CO_2 (>900 W)	—	Yes	—	—
Low/medium-power Nd:YAG (<100 W)	—	—	Yes	—

Figure 13-2 shows one example of a laser system that has been used for a number of years for production welding of precision parts. It is based on a continuous 600-W CO_2 laser, in the large cabinet at the left. The CO_2–N_2–He gas mixture flows through three tubes, each 3 m long and arranged in a Z shape. The operation of the laser is controlled by a programmable digital controller, which turns the laser on and off at prescribed times for automated welding. Safety features include transmission of the beam to the workstation through a metal pipe, and enclosure of the workstation with lucite panels, which are opaque at 10.6 μm. These panels are interlocked with the beam shutter so that the beam cannot enter the workstation when the panels are opened. The beam is turned through 90° by a mirror and focused on the workpiece by a lens. The workpiece is held in a rotary fixture and rotated under the beam to make seam welds. The motion of the workpiece is controlled automatically by the programmable numerical controller. An argon gas jet shields the weld area to exclude nitrogen from the weld area. This example gives an idea of some of the elements that are employed in a complete system for laser welding.

Two considerations are worth emphasizing:

1. The optical path can be completely enclosed. A pipe with prisms to provide bending of the beam can be very effective and can enhance safety, especially with infrared lasers. Several manufacturers offer complete beam delivery systems, including rigid arms, articulated arms, and robot-controlled systems.

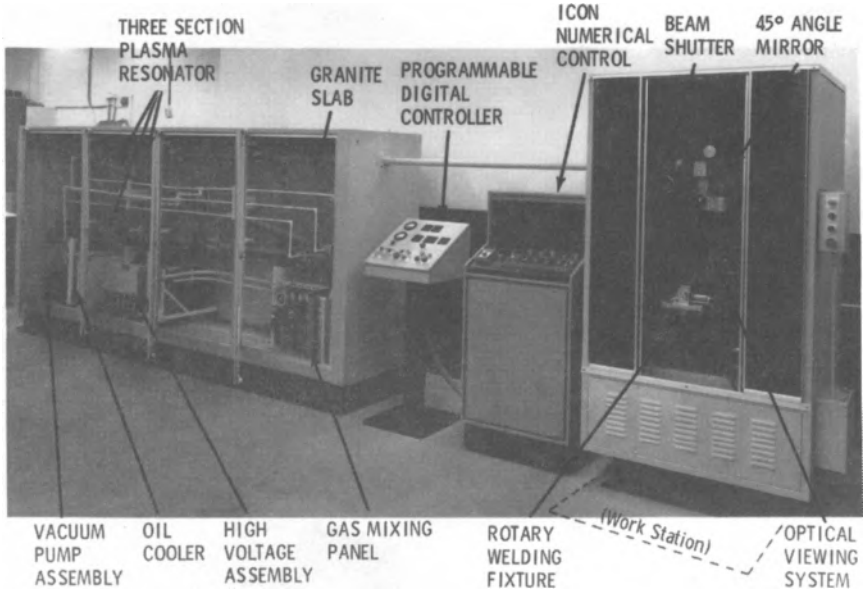


Figure 13-2 Industrial welding system based on a 600-W CO_2 laser.

2. Often, it may be desirable to insert a sheet of transparent material between the focusing optics and the workpiece, to protect the optics from blowoff material. It is much less expensive to replace this disposable material than to replace damaged optics.

A variety of manufacturers from a number of countries offer laser material processing systems. The system designs vary substantially, some incorporating workpiece motion, some incorporating beam motion, and some combining the two. Many systems, especially for sheet metal cutting, employ gantries. Robotic systems offering substantial flexibility are also available. A common feature is the ability to program and control the entire operation on a computer. The user should consider carefully the requirements for a processing operation before selecting a system from the many available.

Specific material processing operations include welding, heat treating, hole drilling, cutting, and scribing. In the next four chapters, we describe some of these applications in more detail. In the remainder of this chapter, we briefly discuss the emerging application of stereolithography, which is somewhat different from the more established material processing applications.

Stereolithography is useful for rapid prototyping. Using stereolithography, one may fabricate a design stored in a computer. The part is fabricated section by section by photopolymerization of a liquid. Stereolithography provides desktop manufacturing of models with fast turnaround and low tooling. It is compatible with computer aided design (CAD).

In perhaps its most common form, stereolithography uses an ultraviolet laser beam, which is scanned across the surface of a curable liquid plastic monomer. In early studies, the ultraviolet line from a helium–cadmium laser (325 nm) frequently was used. The scanning is controlled by a computer to trace a cross section of the desired part. The liquid photopolymerizes and solidifies where it is irradiated. This forms a solid layer, which represents one section of the part. The solidified material rests on an elevator. When one section is completed, the elevator descends by the thickness of one section (perhaps a few thousandths of an inch). Then the laser is scanned in the pattern of the next cross section. This forms the second layer, which adheres to the first. In a number of steps, a model of the entire part is formed. When the model is completed, it may be removed from the liquid bath and cured in an oven.

Stereolithography can take a variety of different forms, including working with powders instead of liquids. The example just given shows the important concepts that apply generally.

Stereolithography allows a computer generated design to be turned into a three-dimensional model very rapidly. Stereolithography can reduce the time required to develop prototype models substantially. Conventional methods could require weeks or months, whereas stereolithography can yield the first model within hours or perhaps minutes after a design is complete. The models produced in this fashion are useful for many purposes, such as verification of design, marketing, testing, and customer preference studies.

This novel process is still in a relatively early stage of development and has not yet reached its full potential. It does offer substantial promise for acceleration of prototype production and for reducing the time that it takes to introduce new products. Some commercial systems for performing stereolithography are becoming available. We may expect this area of laser processing to expand in the future.

Selected References

M. Burns, *Automated Fabrication: Improving Productivity in Manufacturing*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

S.S. Charschan, ed., *Guide to Laser Materials Processing*, Laser Institute of America, Orlando, FL, 1994.

L. Migliore, ed., *Laser Materials Processing*, Marcel Dekker, New York, 1996.

J. F. Ready, Material Processing—an Overview, *Proc. IEEE* **70**, 533 (1982).

C. J. Nonhoff, *Material Processing with Nd Lasers*, Electrochemical Publications, Ltd., Ayr, U.K., 1988.

Chapter 14 | Applications of Laser Welding

In the preceding two chapters, we have described the physical mechanisms by which high-power laser radiation interacts with materials and the potential applications of the laser–surface interactions for material processing. In this chapter, we describe the specific application of welding in greater detail.

Early in the history of lasers, it was recognized that specialized welding applications could be performed with the ruby lasers then available. In certain cases, laser welding provided technical advantages, such as a very small heat-affected zone. Most of the early studies that defined welding capabilities were based on ruby lasers. Most studies emphasized spot welds, but seam welding by overlapping pulses was also demonstrated, although at slow rates. By the late 1960s, some production applications of spot welding with ruby lasers were announced. By 1970, the development of CO_2 and Nd:YAG lasers had made them the leading contenders for welding applications, and they have remained the laser types most often used in welding until the present. Seam welds can be made at reasonably high rates with continuous or repetitively pulsed CO_2 or Nd:YAG lasers. In 1971, deep penetration welding with multikilowatt CO_2 lasers was announced. By the mid 1970s, laser welding had reached production status for a number of applications. Continued laser development has now made laser welding economically competitive with other welding methods. Laser welding is chosen for production not only because it offers technical advantages, but because it costs less in many cases.

The parameters of the laser beam and the properties of the workpiece strongly influence the results of a laser welding application. The thermal diffusivity of the workpiece is important. High thermal diffusivity allows faster conduction of the heat energy through the workpiece and in general permits greater depth of welding. High surface reflectivity can reduce the energy absorbed by the surface. Although surface reflectivity can drop during the interaction (see Chapter 12), welding of metals with high reflectivity does require more energy than welding of low-reflectivity metals. Surface reflectivity may influence the choice of the laser to be used, so that one may

operate at a wavelength where the reflectivity is relatively low. A shiny surface may be darkened by a coating, but this is not always effective; the coating may be rapidly vaporized, leaving the shiny metal unaffected. Surface finish can also influence light absorption and hence the weld penetration. The depth of penetration generally is decreased by polishing a metallic surface.

Many properties of the laser must be properly chosen to optimize the interaction. These include the following:

- The laser wavelength should be absorbed well by the workpiece.
- The irradiance at the surface must be high enough to produce melting.
- For pulsed lasers, the pulse duration must be long enough to permit penetration of the heat energy through the workpiece.
- The pulse repetition rate must be high enough to weld a seam at a reasonably high rate.
- The irradiance and pulse duration must fall into a regime where surface vaporization is not excessive.

As noted before, in welding applications, the material should be melted with penetration of heat through the sample. Vaporization of the surface should be avoided, and the pulse duration or dwell time should be tailored to allow adequate penetration. With good control of the laser pulse, vaporization can be minimized. Optimization of the weld process is largely controlled by the pulse length. For many pulsed welding applications, it has been desirable to stretch the pulse length to allow greater penetration.

Perhaps the most important parameter in laser welding is the irradiance delivered to the surface. This is determined both by the laser output and the focusing. Developments in laser technology are leading to lasers with increased radiance, which translates into higher irradiance at the surface and faster welding rates.

In the setup of a laser welding operation, one cannot regard the laser simply as an individual component; rather, one must treat the entire system as an integrated entity to perform the desired job. The following considerations are important for establishing a successful laser welding application:

- Selection of tasks compatible with laser capabilities
- Economic analysis to determine suitability for laser use
- Choice of proper type of laser
- Determination of optimized process parameters (pulse duration, irradiance, etc.)
- Design of fixturing for the workpiece
- Design of optical system for beam control and delivery
- Development of system to move the workpiece in the beam or to move the beam over the workpiece in a prescribed pattern
- Real-time measurement of the effect on the workpiece, with feedback control of the laser
- Development of appropriate safety measures

Thus, one does not simply aim a laser at a workpiece and shoot. One must instead carefully design an entire integrated system, of which the laser is only one part.

The discussion of laser welding in this chapter will begin with seam welding by lasers with average power in the subkilowatt range, a regime in which penetration of heat by thermal conduction is the dominant factor. Then, we discuss welding with multikilowatt lasers, for which the limitations imposed by thermal conduction are relaxed in a deep penetration mode of welding. We then describe spot welding applications, for which average power is not always an important consideration. Finally, we present two specific examples of the use of laser welding to solve production problems.

A. Seam Welding: Subkilowatt Levels

We describe first seam welding at levels below 1000–1500 W. Seam welding requires either a continuous laser or a repetitively pulsed laser with reasonably high pulse repetition rate, and with at least a few hundred watts average power. Thus, most seam welding applications have employed either CO_2 lasers or Nd:YAG lasers. Development of new lasers with high radiance may change this in the future, however. At average power levels below 1000 W, the irradiance at the surface is limited, and one works in the regime where the laser energy is absorbed at the surface of the workpiece and penetration is limited by thermal conduction. At average power levels of several kilowatts, other physical phenomena become operative. Thus, welding with multikilowatt lasers will be described separately in the next section.

At 100 W average power, a repetitively pulsed laser is a welding tool, at least for some metals. A continuous laser must emit at least several hundred watts to have comparable metalworking capability, because the high peak power in the pulse can break down the surface reflectivity, allowing more efficient absorption of the laser energy. At several hundred watts average power, both the pulsed and the continuous CO_2 lasers are well suited for seam welding. The pulsed CO_2 laser forms a seam by overlapping spots, but the pulse repetition rate is high enough to form seams rapidly. The welding capability of a pulsed CO_2 laser is enhanced, if one pulses the power supply so as to deliver pulses with a higher-power initial spike and a lower-power tail, similar to what is shown in Figure 2-13. The initial spike breaks down the surface reflectivity and allows more efficient coupling of the later portion of the pulse, which contains most of the energy.

Similarly, the repetitively pulsed and continuous Nd:YAG lasers operated at several hundred watts average power are effective welding tools. At equal power levels, the Nd:YAG laser usually has greater welding capability than the CO_2 laser, as we shall see later. But because the development of the CO_2 laser with high-power capability has advanced faster than the Nd:YAG laser, CO_2 lasers have perhaps been employed more often to date in production applications. Advances in solid state laser technology may lead to the use of more solid state lasers in the future.

The trade-off between penetration depth and welding speed for seam welding with a 375-W continuous CO_2 laser is shown in Figure 14-1. Results are shown for welding of two different types of steel. These results are for butt welding, with full penetration of the weld zone through the material. Many potential applications can be satisfied by combinations of penetration and weld speed as shown in the figure. But high weld speeds are possible only with thin material, because of the constraints imposed by thermal conduction. Even at very slow rates, penetration is limited to ~0.04 in. for stainless steel and ~0.06 in. for carbon steel. The lower penetration for stainless steel is a result of its lower thermal conductivity.

Figure 14-2 shows a similar curve for type 302 stainless steel, welded by a 1500-W continuous CO_2 laser. Although the power was in excess of 1000 W, the welding was not of the deep penetration type to be described later. The increased power allowed deeper penetration, but even at slow speed the maximum penetration is limited to ~0.12 in.

It is worthwhile to compare the welding obtained using a CO_2 laser and a Nd:YAG laser. Figure 14-3 shows weld rates obtained with a Nd:YAG laser welding stainless steel, at three different power levels. Comparison of these data with those of Figure 14-1 indicate that at the level of a few hundred watts, the results are qualitatively similar for the two types of lasers. The Nd:YAG laser does offer somewhat increased penetration or weld speed, because of the increased coupling at the shorter wavelength and because of the increased irradiance that arises from finer focusing. The difference in surface reflectivity at 1.06 μm and 10.6 μm is not too important

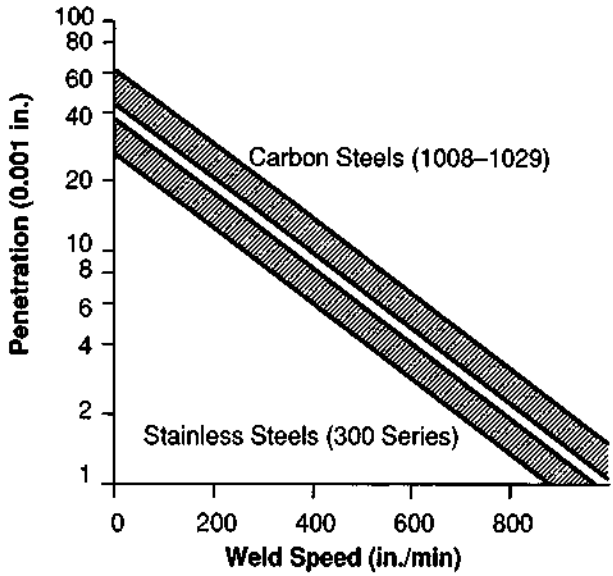


Figure 14-1 Penetration depth versus weld speed for some steels welded with a 375-W continuous CO_2 laser. (Data from Photon Sources, Inc.)

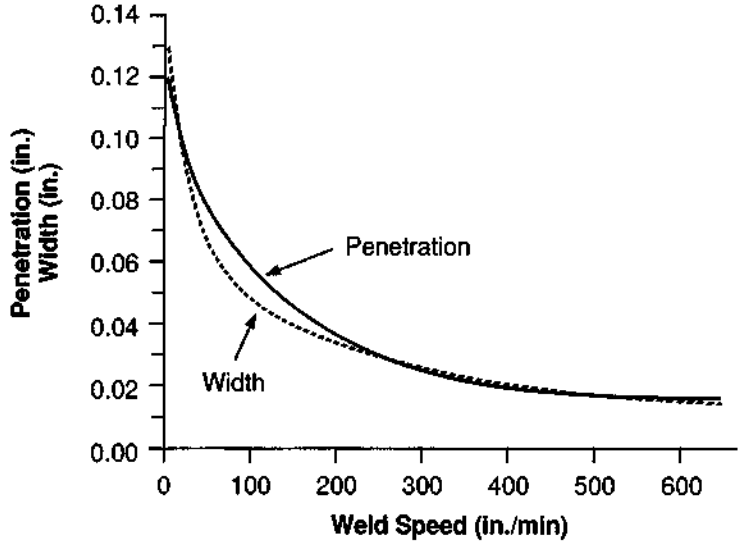


Figure 14-2 Weld rate versus penetration depth and weld width for welding of 302 stainless steel with a 1500-W continuous CO_2 laser. (From S. L. Engel, *Laser Focus*, p. 44 (February 1976).)

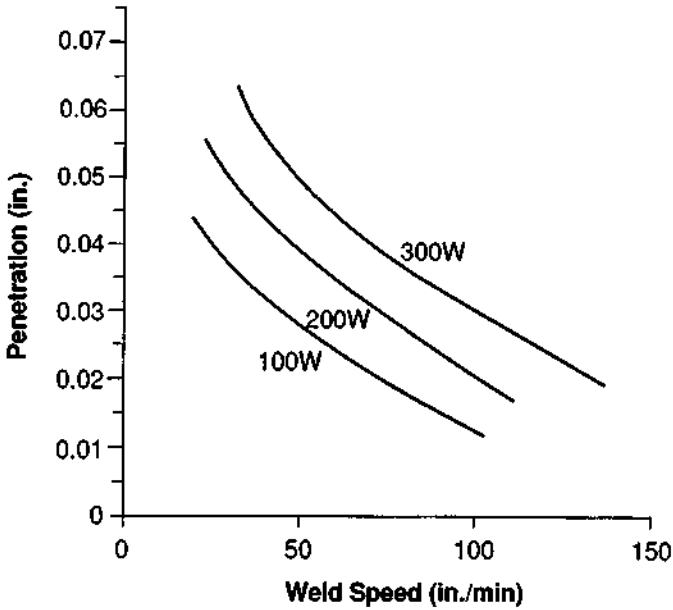


Figure 14-3 Weld penetration as a function of speed for stainless steel with a Nd:YAG laser at the indicated power levels. (From data of Lasag AG.)

for steel. Surface breakdown caused by the laser means that the energy is absorbed reasonably well in either case. But for high-conductivity metals (like Al and Cu, see Figure 12-3) the difference is more significant, and Nd:YAG lasers are better suited for such welding.

Some data on seam welding with a pulsed Nd:YAG laser are shown in Figure 14-4. The results are relevant to a Nd:YAG laser operating at an average power of 400 W, with a peak power of several kilowatts, and operated in several different modes. The material is 300 series stainless steel. The "Conduction Weld" curves are for conditions of focusing such that no substantial amount of vaporization occurred. The "Penetration Weld" curve is for conditions of focusing to higher values of irradiance at the surface, so that substantial vaporization did occur. A hole was produced in the material, and deeper penetration was achieved. This dashed curve represents the onset of deep penetration welding, which will be described in the next section. In the penetration mode, larger values of weld depth become possible. Data are also shown for seam welding with a continuous laser of the same average power. The results for the continuous laser are comparable to those shown in Figure 14-3. The pulsed laser gives deeper penetration because the higher peak power in the pulse yields more efficient coupling of the energy into the sample.

The weld penetrations shown in Figures 14-1 to 14-4 are representative of what can be achieved with lasers emitting some hundreds of watts of average power. In this

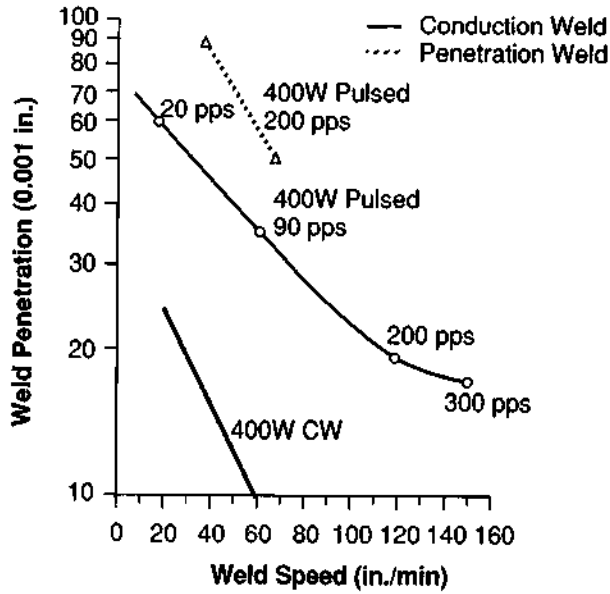


Figure 14-4 Weld penetration versus weld rate for Nd:YAG welding of stainless steel. The figure shows weld penetration for different operating conditions at the same average power. (From H. L. Marshall, in *Industrial Applications of High Power Laser Technology* (J. F. Ready, ed.), *SPIE Proc.*, Vol. 86, 1976.)

regime, the laser energy is absorbed at the surface of the workpiece, and penetration of the energy into the sample is due to thermal conduction. This fact restricts penetration depth to a few hundredths of an inch. The penetration could be increased by long dwell times with a continuous laser. But this would result in large heat-affected zones and slow weld rates. The use of multikilowatt lasers relieves these problems and allows welding with much deeper penetration, as the next section will describe.

The data in Figures 14-1 to 14-4 are drawn from different sources. In comparing data from different sources, some apparent discrepancies may be found. Conditions such as surface finish, beam profile, focusing, irradiance, and the like will vary and make it difficult to intercompare data. Thus, results such as those shown in these figures must be accepted as approximate. The quantities, such as penetration, weld rate, heat affected zone, and so forth, obtained in a material processing study depend strongly on the exact conditions, which vary from one laser installatioon to another. Results shown as examples in this book are intended to provide guidance about general trends and principles. For more exact values, the user's own system must be employed to determine parameters such as penetration depth and weld rate under the exact conditions that are to be used.

The data in Figures 14-1 to 14-4 have all been concerned with steel. Table 14-1 presents statements on the relative ability of lasers to weld a number of common metals [1]. The weld quality is given qualitatively for both CO_2 lasers and Nd:YAG lasers. For ferrous metals (steels, nickel alloys), the CO_2 laser generally performs well. But for high-conductivity metals (copper, silver), the Nd:YAG laser performs better. For alloys with a component with low vaporization temperature (brass), that component may boil away rapidly, so that the resulting weld may be porous.

Laser welding is also applicable to dissimilar metals. Table 14-2 presents some results obtained for the quality of the welds formed by a number of binary metal combinations. Many combinations of dissimilar metals can be welded to produce welds of good or excellent quality. For some other metals, including lead, tin, zinc, and aluminum, the quality of the weld formed with another metal is usually fair or poor. Precious metals, like gold, silver, platinum, and palladium, usually form excellent welds with another metal of that group, but they give mixed results when welded to a metal of another type. Of course, if the basic metallurgy of the alloy formed by the two metals is poor, the weld quality will be poor. Laser welding will not cure poor metallurgical combinations.

Laser welding is complete fusion welding, with melting and intermixing of the constituent pieces. This is illustrated in Figure 14-5, which shows a longitudinal section along the top of a lamination stack of high cobalt alloy plates, with one stainless steel end plate. The flow and mixing of the molten material during the welding process is apparent. Also apparent is the smallness of the heat-affected zone next to the weld.

Laser welding can be full strength welding. The tensile strength of the material in the weld region can be as high as that of the original material.

It is worthwhile to consider some of the practical aspects of producing high-quality laser welds. To achieve good welding, the two pieces to be welded must be fitted closely together, so that the materials from each part can intermix well. Generally, the beam should be delivered approximately equally to each piece.

Table 14-1 **Weldability of Various Metals**[a]

Metal	CO_2 laser weldability	Nd:YAG laser weldability	Comments
Aluminum	Fair–good	Fair–good	High reflectivity at 10 μm, some alloys may require filler
Beryllium–copper	Poor–fair	Good	
Brass	Poor	Poor–fair	Zinc boiloff may cause porosity
Carbon steel	Excellent	Fair–good	
Copper	Fair–good	Good–excellent	High reflectivity at 10 μm
Galvanized steel	Fair–good	—	Zinc boiloff may cause porosity
Hastalloy X	Poor–fair	Good–excellent	Need high rep rate to prevent cracking
Inconel	Excellent	Fair–excellent	
Monel	Good	Good	
Nickel	Good	Good	
Silver	—	Excellent	Too high reflectivity at 10 μm
Stainless steel	Good–excellent	Good–excellent	
Titanium	Good–excellent	Good	
Zircalloy	Good	—	

[a]Adapted from *The Industrial Laser Annual Handbook—1990 Edition* (D. Belforte and M. Levitt, eds.), PennWell Books, Tulsa, OK, 1990.

Table 14-2 **Laser Weldability of Binary Metal Combinations**[a]

	Ta	Mo	Cr	Co	Ti	Fe	Ni
W	Excellent	Excellent	Excellent	Fair	Fair	Fair	Fair
Ta		Excellent	Poor	Poor	Excellent	Fair	Good
Mo			Excellent	Fair	Excellent	Good	Fair
Cr				Good	Good	Excellent	Good
Co					Fair	Excellent	Excellent
Ti						Fair	Fair
Fe							Good

[a]Adapted from *The Industrial Laser Annual Handbook—1990 Edition* (D. Belforte and M. Levitt, eds.), PennWell Books, Tulsa, OK, 1990.

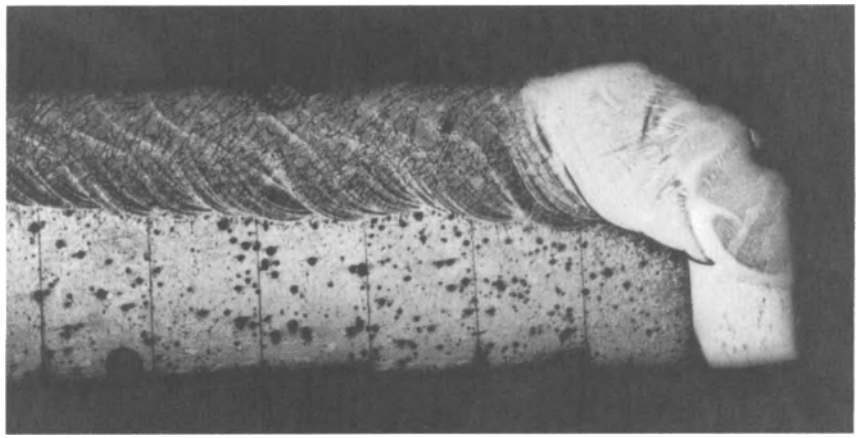


Figure 14-5 Section of a weld along the top of a stack of plates of high cobalt alloy made with a 500-W continuous CO_2 laser. The right end plate is stainless steel. The width of the area shown is 0.1 in. (Photograph courtesy of P. E. Gustafson.)

The workpiece must be positioned and held accurately. This operation is called fixturing. The possible types of fixturing vary widely depending on the exact size and shape of the workpiece. For example, if two pieces of metal are to be butt welded, they must be positioned and held closely together. Then they are moved (or the beam is moved) so that the interface between the pieces passes through the focus of the beam. It is not possible to discuss all possible types of fixturing. We shall simply state that proper fixturing is essential for success of a laser welding operation.

Some rules of thumb for fit-up tolerances for a butt weld are shown in Figure 14-6. The tolerances are stated in terms of material thickness t. The parts should be positioned with a gap no larger than $0.15t$, and the misalignment should be less than $0.25t$. The edges should be square and straight. Compressive clamping is desirable. Adherence to these restrictions aids the formation of a fully welded butt joint. The figure also shows tolerances for a lap joint.

We should also consider the effect of varying the focus of the laser beam. Figure 14-7 shows how the penetration depth and bead width vary for welding of 1018 steel. The welding was done with a 1500-W continuous CO_2 laser at a speed of 50 in./min. The effect of changing focus is apparent, with maximum penetration occurring when the beam is focused at a point slightly below the surface. When a shorter focal length lens is used, the penetration is deeper, because the irradiance is increased as a result of the finer focusing of the beam (see Chapter 2).

The smallness of the heat-affected zone is one of the attractive features of laser welding. Because the heat energy must be conducted through the material, it is difficult to make laser welds that are narrower than the material thickness. In thin materials, seams with widths of a few thousandths of an inch are possible.

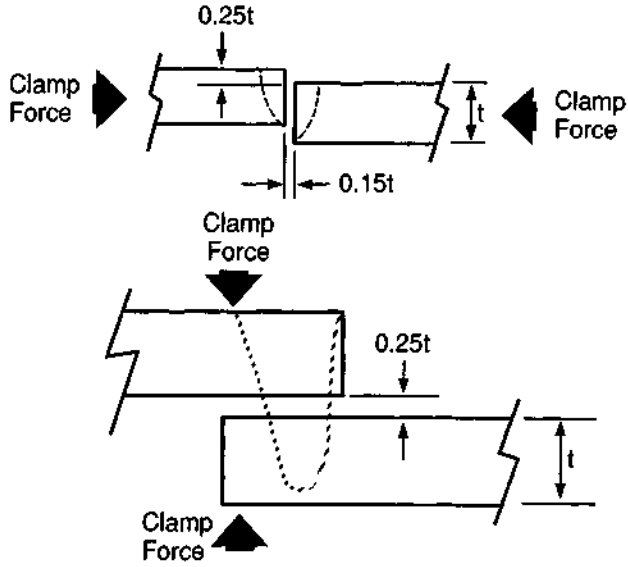


Figure 14-6 Tolerances for butt welding (top) and for lap welding (bottom) for material of thickness *t*. (From S. L. Engel, *Laser Focus*, p. 44 (February 1976).)

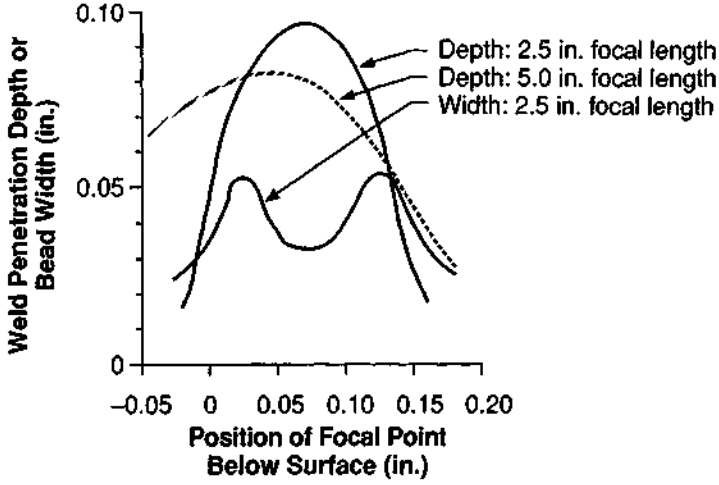


Figure 14-7 Weld penetration and weld width as a function of focal position for welding of 1018 steel at a speed of 50 in./min with a 1500-W continuous CO_2 laser, with lenses of the indicated focal length. (From S. L. Engel, *Laser Focus*, p. 44 (February 1976).)

There have been many successful applications of laser welding with CO_2 lasers or Nd:YAG lasers with average power of several hundred watts. The applications have included hermetic sealing of gyroscope cases, welding of caps to vanes in gas turbine engines, formation of contour welds on inconel turbine blades, welding of end caps on zircalloy nuclear fuel rods, and welding of hardened steel edges to steel band saw blades for meat cutting.

In the automotive industry, lasers have been used for many operations, including welding of bearing race retainers, throttle bodies, and heater–motor yoke assemblies. The automotive industry also uses lasers for welding of automatic transmission and air bag components. These devices require low heat input and low distortion. As of the mid 1990s, automobile body welding is not widely performed by lasers, but there are predictions that by the early twenty-first century laser welding of automobile bodies will become significant.

One example of a very successful laser welding application is the welding of window spacers, used between sheets of glass in insulated windows. Flat aluminum stock is roll formed into a tubular shape. The joint of the tube is then welded with a CO_2 laser beam, typically with a power around 1500 W. The laser welding process produces high-quality welds with good cosmetic appearance. The use of laser welding for this application has lowered production costs and provided material savings compared with conventional techniques.

B. Welding with Multikilowatt Lasers

We have so far emphasized the limited penetration capabilities for seam welding by lasers with average power below the 1000–1500 W level. With the development of high-power continuous CO_2 lasers around 1970, deeper penetration welding became possible. Figure 14-8 shows penetration depth as a function of laser power. The dashed curve is for welding at levels below 1000 W; as noted before, penetration is limited. Around 1000–1500 W, new physical phenomena become important, and the curve bends upward. Curve *A* represents an early correlation of data for a variety of optical beam parameters and for several different metals. This correlation indicated that the penetration increased as the 0.7 power of the laser output, reaching almost 2 in. for steel at the 80 kW level. Curve *B* represents a more recent compilation, for beams with low divergence. This curve indicates an increase of penetration as the 0.8 power of laser output, with the penetration reaching 1 in. near 11 kW.

In deep penetration welding, the beam delivers energy to the surface rapidly. An irradiance of 10^6 W/cm^2 or more is required. If this condition is satisfied, a hole is drilled into the material and laser energy is deposited throughout the depth of the hole. Thus, much deeper penetration is possible, because one does not rely on thermal conduction of energy deposited at the surface. This process, which is illustrated in Figure 14-9, is called "keyholing." The figure shows the formation of the keyhole.

As the laser beam moves across the surface, the hole is translated through the material. Surface tension causes the molten metal to flow around the hole and fill in behind it. The molten metal then rapidly resolidifies, forming a seam weld.

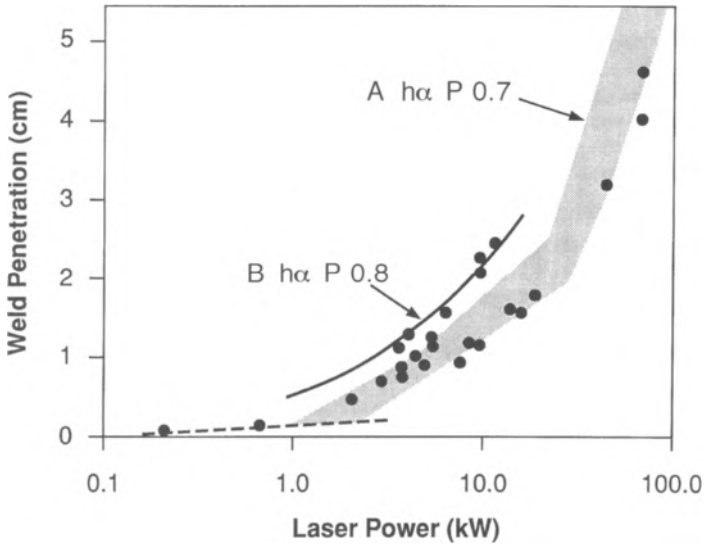


Figure 14-8 Weld penetration as a function of laser output power. Below 1 kW, the process is dominated by conduction; above 1 kW, deep penetration welding becomes possible. Curve *A* represents data from a number of sources, and it indicates that the penetration depth h increases as the 0.7 power of the laser power P. Curve *B* represents more recent data obtained with a laser with improved beam focusability; it indicates that h scales as the 0.8 power of P. (From C. Banas, in *The Industrial Laser Annual Handbook—1986 Edition* (D. Belforte and M. Levitt, eds.), PennWell Books, Tulsa, OK (1986).)

At the irradiance required for keyholing, a plasma may be formed above the surface, with resultant beam attenuation. Thus, in deep penetration welding applications, one often uses a jet of high-velocity gas to remove the plasma.

The keyholing process is very similar to what occurs in electron beam welding. Like an electron beam, a multikilowatt laser can produce narrow, deep welds. But unlike an electron beam, the laser beam may be transmitted through the atmosphere, and one may use simplified tooling like vacuum chucks. We shall return to a more quantitative comparison of laser and electron beam welding later.

Figure 14-10 shows cross sections of welds from bead-on-plate welding of 304 stainless steel at different speeds. These results were obtained with a 15-kW CO_2 laser. At low speed, welds with penetration of 0.8 in. were attained in a single pass. As speed increased, the weld bead became narrower, but the penetration decreased slowly, still exceeding 0.6 in. at a speed of 60 in./min. The shapes of the weld nuggets illustrated in the figure are characteristic of multikilowatt deep penetration laser welding. The shape of the weld zone is determined by a number of factors. One controllable factor is the location of the focus relative to the surface of the workpiece. In one experiment using a 16-kW CO_2 laser [2], it was found that maximum penetration occurred when the beam was focused slightly inside the workpiece.

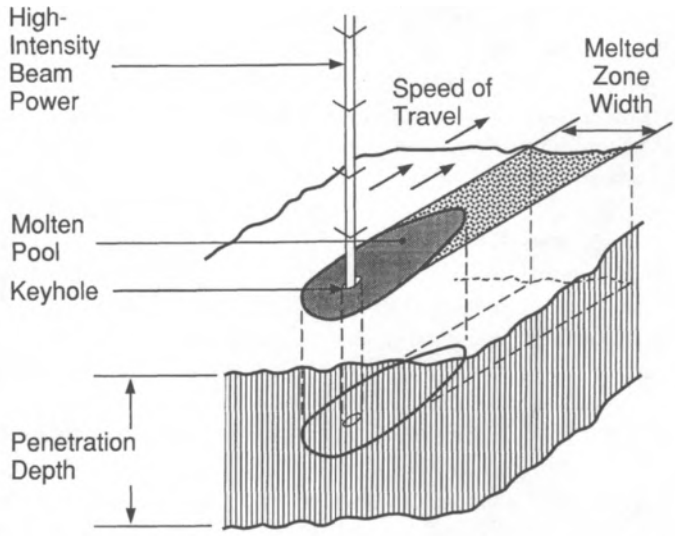


Figure 14-9 The process of keyhole formation and deep penetration welding. (From D. T. Swift-Hook and A. E. F. Gick, Welding Research Supplement, *Weld. J.*, p. 492-S (November 1973).)

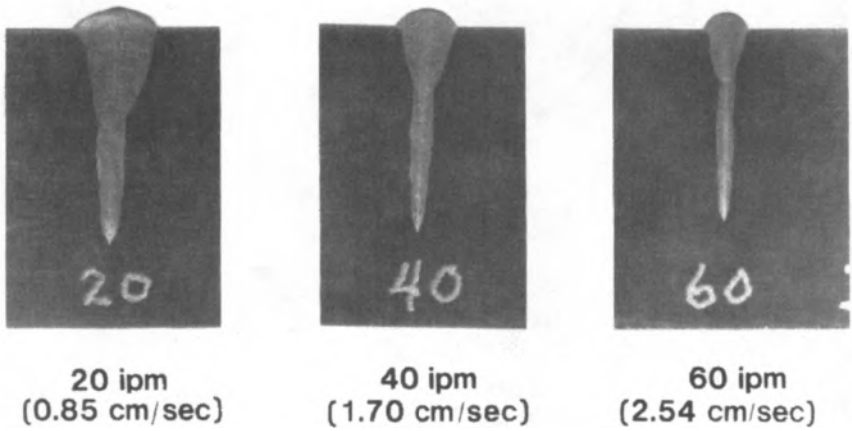


Figure 14-10 Weld beads produced in 1-in.-thick plates of 304 stainless steel in bead-on-plate welding with a 15-kW CO_2 laser, at the indicated values of welding speed. (From E. M. Breinan, C. M. Banas, and M. A. Greenfield, United Technologies Research Center Report R75-111087-3, 1975.)

Penetration decreased when the beam was focused at the surface or deeper inside the workpiece. Some cross sections of the weld zone had an hourglass shape, with the position of the waist of the hourglass being controlled by the position of the focus.

The deep penetration capability of multikilowatt laser welding is illustrated in Figure 14-11. The figure summarizes data on the penetration achieved in 304 stainless steel at various welding speeds, for laser power up to 17 kW. The penetration depths and welding speeds are large enough to allow great capability for practical welding.

The results in Figure 14-11 are representative of some of the many data presented in the literature for deep penetration welding. The results obtained in a particular welding application will depend on the specific circumstances, including focusing, the velocity, type and angle of the gas jet used to dissipate the plasma, and so forth. Data from the literature have been compiled by V. G. Gregson in the form of an energy boundary that defines the limitations of the welding process. An example is shown in Figure 14-12, for bead-on-plate welding of mild steel. This figure shows the maximum penetration depth attainable as a function of energy per unit length (defined as the power divided by the weld speed). It represents the deepest penetration possible for current technology, not necessarily optimum welding conditions. High-quality welds may be produced in the region below the boundary. The discontinuity in the curve near 2 kJ/cm may represent the transition from conduction welding to deep penetration keyhole welding.

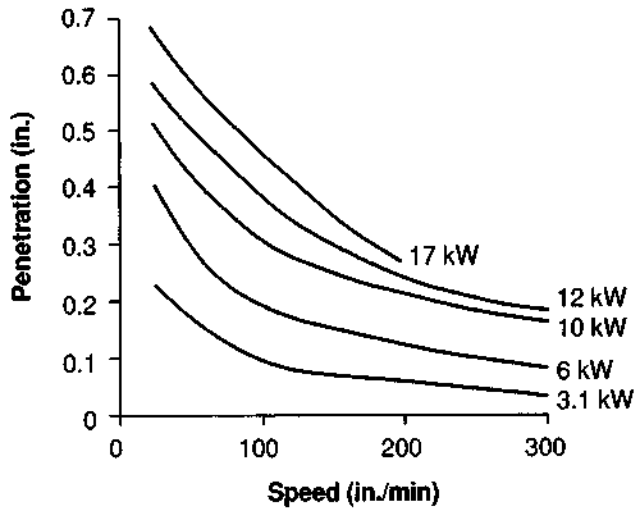


Figure 14-11 Weld penetration as a function of weld speed for CO_2 laser welding of 304 stainless steel at the indicated values of laser power. (Adapted from E. V. Locke and R. A. Hella, *IEEE J. Quantum Electron.* **QE-10**, 179 (1974).)

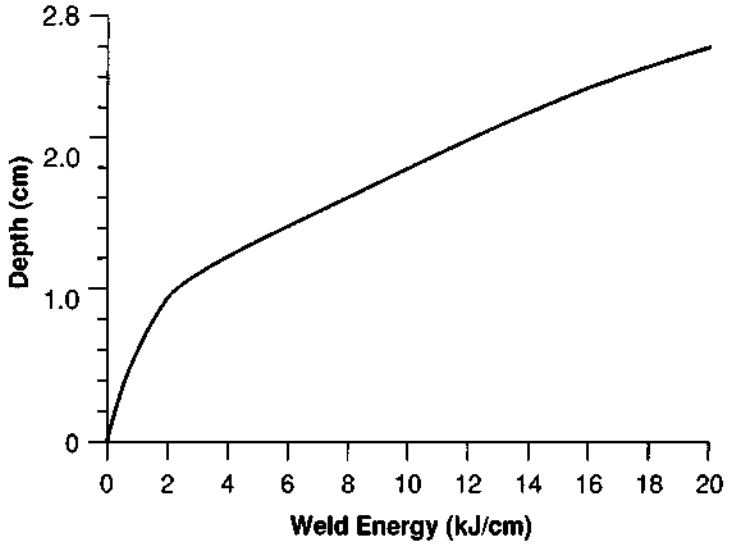


Figure 14-12 Maximum penetration depth as a function of energy per unit weld length for bead-on-plate welding of mild steel. (This curve is based on data compiled by V. G. Gregson, Westinghouse Marine Division.)

Deep penetration laser welding has been performed mostly with multikilowatt CO_2 lasers. Until recently, there have been no well-developed multikilowatt Nd:YAG lasers. This situation has changed, and reliable stable Nd:YAG lasers with average power capability up to 1800 W are now available. The deep penetration welding capabilities of such a laser are shown in Figure 14-13, which shows results for penetration depth as a function of welding speed at various power levels for welding of 304 stainless steel. At the 1800 W level, the penetration approaches 1 in. A comparison of this figure with Figure 14-11 indicates that a suitable Nd:YAG laser can perform penetration welding at lower levels of power than a CO_2 laser, largely because of the higher radiance, which means that a higher value of irradiance may be delivered to the surface of the workpiece. The use of high-radiance lasers offers possibilities for expanded laser welding capabilities. We shall return to this point later.

Multikilowatt laser welding competes for applications with electron beam welding; the applications for either type of welding tend to be similar. Most comparisons between laser welding and electron beam welding indicate that electron beam welding has somewhat greater capabilities than laser welding. Figure 14-14 shows an early comparison between laser welding and hard-vacuum electron beam welding at the 10 kW level for 304 stainless steel. At speeds from 50 to 500 in./min, the laser penetration is approximately 70 percent of the electron beam penetration. The electron beam penetration continually increases as speed decreases, whereas penetration by the laser beam appears to saturate. This behavior may be associated with the production of a plasma.

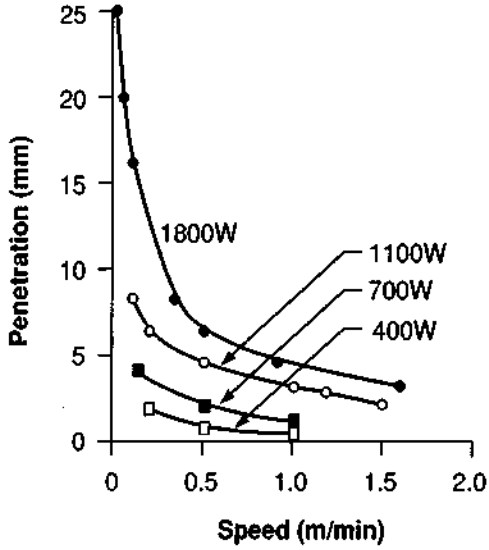


Figure 14-13 Weld penetration versus welding speed for welding of 304 stainless steel by pulsed Nd:YAG lasers delivering to the workpiece average powers from 400 to 1800 W in pulses with 0.6 msec duration at a 100 Hz pulse repetition rate. (From C. L. M. Ireland, *Laser Focus/Electro-Optics*, p. 49 (November 1988).)

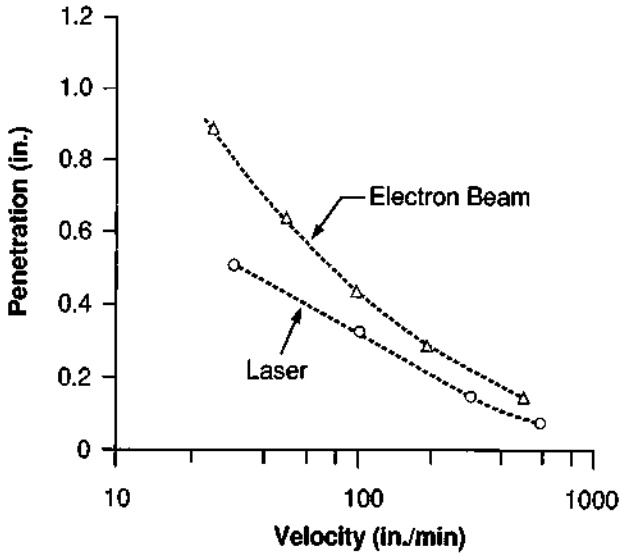


Figure 14-14 Comparison between laser welding and vacuum electron beam welding of 304 stainless steel at the 10 kW level. The curves show penetration depth as a function of weld velocity for the two processes. (From E. V. Locke and R. A. Hella, *IEEE J. Quantum Electron.* **QE-10**, 179 (1974).)

Laser welding, though, offers factors that may offset this apparent advantage for the electron beam. Laser welding can be performed without the need for any vacuum system and thus can offer a higher throughput of completed parts per unit time than the electron beam. Even though the electron beam can weld an individual part faster than the laser, the necessity of moving the parts into and out of a vacuum can slow the electron beam processing to a rate below that of the laser processing. The fixturing requirements also are simpler for laser welding. In fact, laser welding has replaced electron beam welding in a number of applications. The higher throughput allows lower unit cost per welded part.

In addition, more recent work with high-radiance lasers [3] has indicated that laser welding can be performed with efficiency comparable to that of electron beam welding. Thus, as high-power lasers with high radiance become more common, we may expect laser welding to further displace electron beam welding in selected applications.

Deep penetration welding with efficient use of the laser energy is possible. Figure 14-15 shows results for welding of several metals by a 5-kW CO_2 laser. The results are expressed in terms of a dimensionless weld speed parameter, Vb/k, where V is the seam welding rate, b the width of the fusion zone, and k the thermal diffusivity averaged over temperature. The power is also expressed as a dimensionless parameter, $P/hK\Delta$, with P the laser power, h the weld penetration, K the thermal conductivity averaged over temperature, and

$$\Delta = T_m - T_0 + \frac{H}{c} \tag{14.1}$$

with T_m the melting temperature, T_0 the ambient temperature, H the latent heat of fusion, and c the heat capacity. The diagonal lines in the figure are lines of constant melting efficiency, which is defined as the ratio of the energy required to just melt the specified volume of material divided by the laser energy. Because Vbh is the volume of material welded per unit time and $\rho[c(T_m - T_0) + H]$ is the energy required per unit volume to melt the material, then at 100 percent melting efficiency we have

$$\begin{aligned} P &= (Vbh)\rho[c(T_m - T_0) + H] \\ &= \frac{Vb\rho c}{K} hK\left(T_m - T_0 + \frac{H}{c}\right) \\ &= \frac{Vb}{k} hK\left(T_m - T_0 + \frac{H}{c}\right) = \frac{Vb}{k} hK\Delta \end{aligned} \tag{14.2}$$

or, for 100 percent melting efficiency,

$$\frac{Vb}{k} = \frac{P}{hK\Delta} \tag{14.3}$$

This is the line marked 100 percent melting efficiency in the figure. The foregoing relation states that all incident laser energy is used to melt metal, with no losses. Other diagonal lines define particular lower values of melting efficiency.

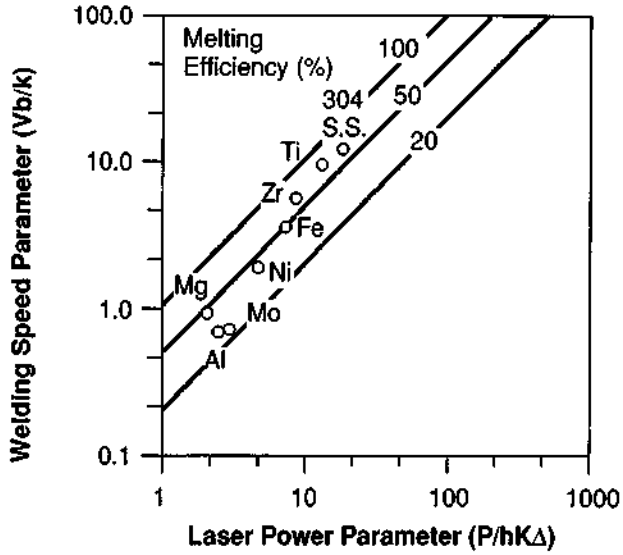


Figure 14-15 Normalized weld speed as a function of normalized laser power. The symbols are defined in the text. The diagonal lines indicate values of the melting efficiency as indicated. (From E. M. Breinan, C. M. Banas, and M. A. Greenfield, United Technologies Research Center Report R75-111087-3, 1975.)

For several metals, especially metals with low thermal conductivity like stainless steel, high values of melting efficiency are attainable, up to 70 percent. This means that 70 percent of the incident laser energy is utilized in melting metal and only 30 percent is lost. Some amount of loss is unavoidable because of thermal conduction out of the weld zone, but the loss can be relatively low. The melting efficiency for metals with higher values of thermal conductivity, like aluminum, molybdenum, and magnesium, is lower, because there is greater conduction of heat out of the weld zone.

The results shown in Figure 14-15 are attainable only if the initial surface reflectivity is reduced during the welding process, as we have described in Chapter 12. This leads to the definition of another measure of efficiency, the energy transfer efficiency. This is defined as the fraction of the laser energy that is actually transferred to the material. Figure 14-16 shows some results, using the same dimensionless parameters as Figure 14-15. The figure compares experimental data points to theoretical results, indicated by the solid curves. The energy transfer efficiency can be as high as 90 percent under suitable conditions. The energy transfer efficiency increases as the laser power increases. It increases if one goes to a shorter wavelength, and it increases if a pulsed laser is used instead of a continuous laser of the same average power. High values of energy transfer efficiency are associated with the phenomenon of keyholing, in which the incident energy is trapped within the keyhole and eventually absorbed.

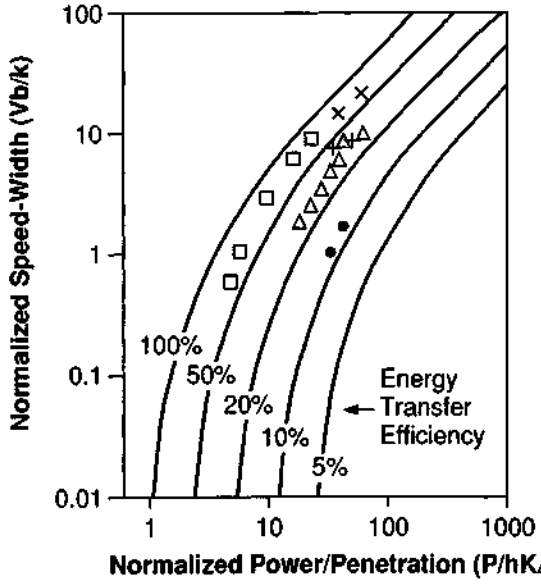


Figure 14-16 Normalized weld speed as a function of normalized laser power divided by penetration depth. The symbols are defined in the text. The solid curves represent constant values of energy transfer efficiency. They are theoretical results, based of the analysis of D. T. Swift-Hook and A. E. F. Gick, Welding Research Supplement, *The Welding Journal*, p. 492-s (November 1973). The experimental data points are the results of C. L. M. Ireland, *Laser Focus/Electro-Optics*, p. 49 (November 1988). This particular figure was developed by S. Hargrove of Lawrence Livermore National Laboratory.

The fact that these measures of efficiency can be high means that the input energy per unit volume required to make a weld is lower than for most conventional welding processes. In fact, it is about a factor of ten lower than for processes like resistance welding. This is one of the attractive features of laser welding; it contributes to the smallness of the heat-affected zone.

It was stated before that laser welds can be full strength welds, with ultimate tensile strength equal to that of the parent metal. Some results are presented it Table 14-3 for laser-welded nickel and iron-based high-temperature alloys [4]. The ductility of the welds was also satisfactory.

Examples of cross sections of welds in Arctic pipeline steel are shown in Figure 14-17. The welds were made with 12 kW of CO_2 laser power, at 25 in./min for the single pass and at 60 in./min for the dual pass. The shape of these weld nuggets is typical of what is obtained with multikilowatt laser welding. They show characteristic columnar grains growing in the direction of the maximum thermal gradient, and they indicate a very small heat-affected zone. Cross-weld tensile tests on these welds resulted in failure of the base metal rather than the weld.

Table 14-3 **Ultimate Tensile Strength in Laser-Welded High-Temperature Alloys**

Alloy designation	Base metal	Thickness (mm)	Ultimate tensile strength in laser weld (Klb/in.^{2})	Ultimate tensile strength in original metal (Klb/in.^{2})
PK33	Ni	1	142	147
PK33	Ni	2	155	151
C263	Ni	1	145	141
C263	Ni	2	141	143
N75	Ni	1	104	117
N75	Ni	2	106	117
M152	Fe	1	129	133
M152	Fe	2	126	125

Commercial lasers are available with outputs up to 25 kW for carbon dioxide lasers and up to 1.8 kW for Nd:YAG lasers. Most of the applications for multikilowatt laser welding have so far used CO_2 lasers, because these were developed first. Because the Nd:YAG laser has greater welding capability at a given value of power, we may expect to see applications for that laser develop.

The first example in which a multikilowatt laser reached production status was in pure lead welding, for production of lead–acid batteries for telephone central offices [5]. The positive plates are cast from lead in a circular assembly with connecting tabs around the diameter. Then tabs are seam welded by a 2-kW CO_2 laser beam

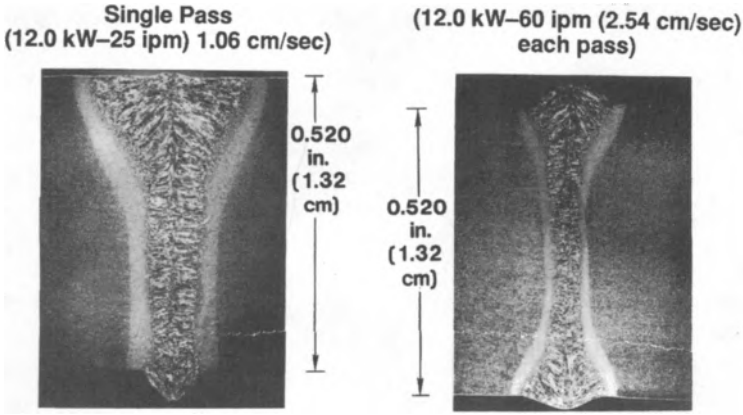


Figure 14-17 Cross sections of laser welds in X-80 Arctic pipeline steel made under the indicated conditions. (From E. M. Breinan, C. M. Banas, and M. A. Greenfield, United Technologies Research Center Report R75-111087-3, 1975.)

moving at 8.5 cm/sec. The laser welder is capable of completing about nine batteries per hour, each with a total of 106 welds. The laser welder replaced four conventional bonders, each of which could produce nine or ten batteries per shift.

Applications of multikilowatt laser welding have become common in the automotive industry. Most of the major automobile manufacturers throughout the world employ numerous lasers. Multikilowatt lasers are used to butt weld sheets of metal to form a larger sheet, which is then stamped to form components like door panels or floor panels in a single operation. Multikilowatt lasers are frequently used for the welding of transmission components. In many of these automotive applications, they have displaced electron beam welding, because the fixturing is easier and there is no need to move the parts into vacuum.

Another developing application of laser welding is the welding of tailored die blanks. This involves the fabrication of blanks of sheet metal of specified shape to be used later in metal forming applications. The blanks for the metal forming operation are produced by laser welding of various flat metal sheets to form a tailored die blank. This welding is usually performed with multikilowatt lasers.

The fabrication of tailored blanks in this way offers several advantages. The physical and chemical properties of the blank can be varied with position in order to influence the deformation process and the final part shape. Factors such as hardness, ductility, thickness, and the presence of coatings can be changed from one location to another within the formed part to meet design requirements. The use of laser die blank welding for industrial applications is increasing, especially in the automotive industry.

Experimental welding has been examined with continuous CO_2 laser powers as high as 100 kW [6]. Stainless steel of 2 in. thickness could be welded at 63 in./min with 80 kW. The use of such high powers for practical industrial welding still requires additional development.

C. Spot Welding

Laser spot welding is a versatile technique for welding of small components, such as attaching wires to terminals. It is particularly useful in cases where localized heating is desired, such as welding near glass-to-metal seals or connection of leads to delicate, heat-sensitive semiconductor circuitry. Good metallurgical joints with small heat-affected zones are produced in this process.

Many of the early studies of laser welding emphasized spot welding, because the only high-power lasers available were ruby and Nd:glass lasers with low pulse repetition rates. Such lasers are best suited for spot welding, and there are many papers in the literature describing such applications. Seam welding was possible by overlapping spots, but this usually was impractical because of the slow seam rate.

More recently, many spot welding applications have been dominated by Nd:YAG lasers, which also have seam welding capabilities, and which can offer higher throughput than a ruby or Nd:glass laser. Ruby and Nd:glass lasers are used less frequently, usually in cases where single pulses with relatively large energy are

needed. A typical application might involve initial spot welding to tack a cover on a package so that it will not move, and then sealing the package with a seam weld. The Nd:YAG laser is well suited for the entire process.

Because much of the data on spot welding was developed with ruby or Nd:glass lasers, we will present some of these results here, because they do define approximate ranges of laser parameters useful for spot welding. Figure 14-18 shows data on the output energy required from a ruby laser to weld copper and nickel wire, as a function of wire diameter. These data are for ruby laser pulses with duration around 3 msec, focused with simple lenses. Lasers with shorter pulse durations (around 1 msec) are better suited for drilling than for welding. The results shown in the figure represent full strength welds: the strength of the weld equals the strength of the original wire. The lines in the figure show the energy required for optimum welding. If the energy is lower, the wire will not be completely melted. If the energy is higher, the wire will be partly vaporized, and the strength will be reduced. More energy is required for copper than for nickel because the reflectivity of copper is higher.

The energy required to penetrate a sheet of either of two different metals when the sheet is irradiated by ruby laser pulses with diameter of 0.4 mm is shown in Figure 14-19. The figure presents the minimum pulse energy necessary for complete penetration of the sheet by the melt puddle as a function of pulse duration for various sheet thicknesses. The figure may be regarded as empirical data on the energy required to spot weld metal of the given thickness. For different beam diameters, the energy should scale approximately as the area of the irradiated spot.

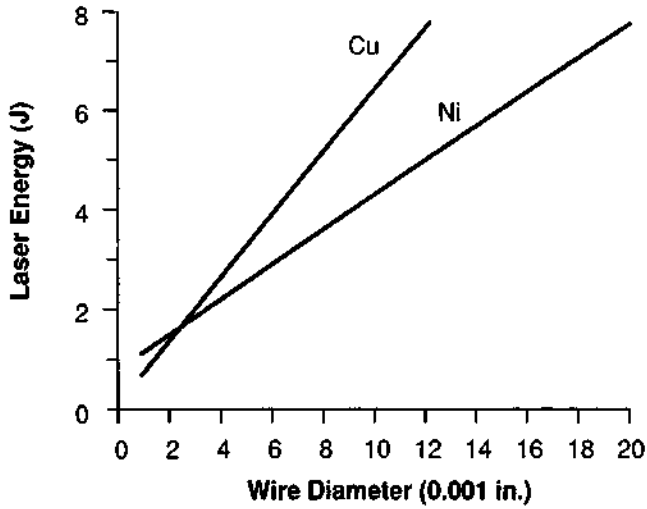


Figure 14-18 Ruby laser energy required to weld copper and nickel wires, as a function of wire diameter, for a 3-msec-duration pulse. (Prepared from data presented by W. A. Murray, *Laser Focus*, p. 42 (March 1970).)

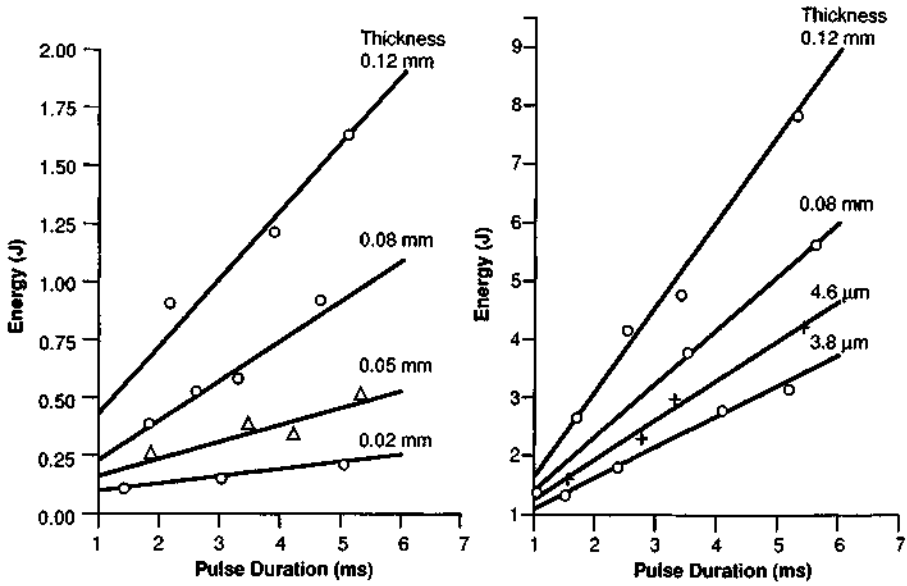


Figure 14-19 Energy required for complete penetration of a metal sheet by a ruby laser pulse with diameter 0.4 mm, as a function of the pulse duration, with the sheet thickness as a parameter. The left portion of the figure is relevant to nickel, the right portion to copper. (From M. I. Cohen and J. P. Epperson, Application of Lasers to Microelectronic Fabrication, in *Electron Beam and Laser Beam Technology*, Advances in Electronics and Electron Physics (L. Marton and A. B. El-Kareh, eds.), Academic Press, New York, 1968.)

Another illustrative application of laser spot welding involves use of a Nd:YAG laser with high pulse repetition rate to weld insulated copper wires to terminal posts [7]. This application had unusual features that illustrate the capabilities of laser welding for practical production situations. First, the laser radiation was used both to remove the insulation and to produce the weld. Second, the weld joints were evaluated in real time by monitoring stress wave emission.

The copper wires were 312 μm in diameter with polyurethane insulation. They were wound on a terminal shoulder that protected them from direct irradiation. This geometry eliminated necking of the wire in the welding process. The laser beam was focused on the terminal head, causing the head to melt and weld to the wire. The laser pulses were 3.5 msec long, delivered 8–10 J of energy, and were focused to a spot size of 0.125 cm on the terminal. The monel terminal post had higher vaporization temperature than copper. When the monel in contact with the wire melted, the copper wire also melted and fused with the monel.

The welds could be evaluated during formation by monitoring stress wave emission, which occurred because of the deformation of material during the welding. The waves were detected with piezoelectric crystals located at a distance from the weld. An empirical relation was developed between the characteristics of the waves

and the weld strength. Characteristic signatures were observed for failure mechanisms like insulation burning and misalignment of the beam. It was possible to identify acceptable welds in real time through analysis of the wave packet associated with the stress waves.

One outstanding application of laser spot welding utilizes Nd:YAG lasers to weld razor blades to a blade support. The application was developed by a major razor manufacturer. Each of two ultrathin stainless steel blades in a twin-blade cartridge receives 13 spot welds. Each 250-W average power Nd:YAG laser makes 1600 welds per minute. The laser pulses are delivered to the workpieces from the remotely located lasers by fiber optic delivery systems. The use of laser welding has been regarded as critical for reducing the costs of this fabrication process.

We will consider two additional illustrative examples of laser spot welding that have reached production status. These examples may give insight into considerations for the practical use of lasers. In both cases, laser welding was able to solve problems that existed with conventional methods.

In the first example [8], a ruby laser beam welded a nickel tab to a nickel alloy post on a transistor. The weld was made with a pulse 3 msec long containing 7.5 J of energy. Production rates of six transistors per minute were attained. The laser welding eliminated problems with cracking at a nearby glass-to-metal seal. With conventional tweezer resistance welding, the heat often caused cracking, and the unit would fail in vibration tests. The laser weld had good intermixing of the constituents, so that the strength was high. The cracking problem was overcome because of the small heat-affected zone, so that good reliability was achieved in vibration tests.

In the second example, two precious metal (Paliney 7) wires were welded to phosphor bronze spring members. The weld is difficult to make with resistance welding because of the difference in electrical resistance of the components. The weld could be made with a 6-J, 3-msec-long ruby laser pulse, focused with a 43 mm focal length lens. The workpiece was 0.075 in. from the focal plane, allowing enough defocusing so that the laser pulse covered both wires and made both welds simultaneously. The heat-affected zone was 0.002 in. wide, so that the temper of the phosphor bronze spring was not affected. The weld had strength at least as great as that of the original wires. Figure 14-20 shows a photograph of the weld.

These examples were chosen to illustrate the capability of laser spot welding for solving difficult technical problems. But laser spot welding can be economical also, displacing competing techniques on the basis of cost. Factors that help reduce the cost of laser welding include high reliability, increased yield, simplified fixturing, and ease of automation, leading to high throughput.

As an example, one analysis [9] compared the cost of laser welding with that of resistance welding. Despite the higher initial cost of the laser equipment, the analysis indicated that over a three year period, making 10 million spot welds per year, the cost per unit for the laser welding was less than 75 percent of the resistance welding. In addition, the laser welding improved weld quality and eliminated thermal damage and part distortion. This example shows that laser spot welding can

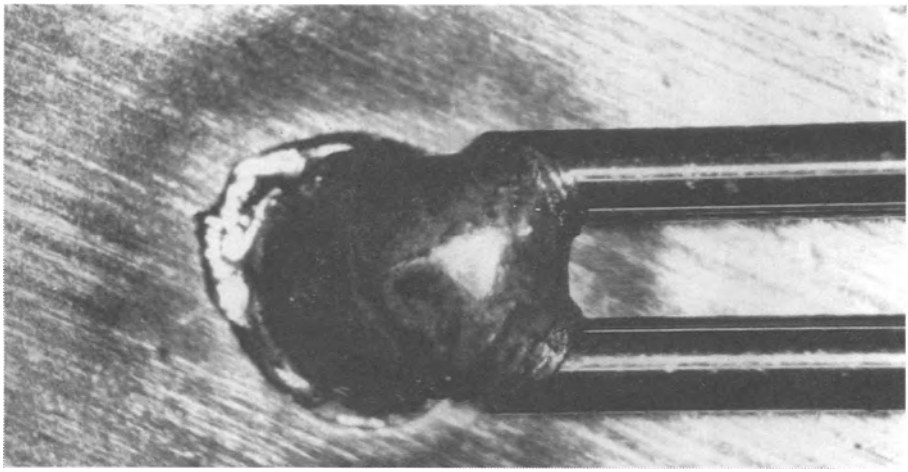


Figure 14-20 Paliney 7 wires welded to a phosphor bronze spring by a 6-J ruby laser pulse. (From F. P. Gagliano, R. M. Lumley, and L. S. Watkins, *Proc. IEEE* **57**, 114 (1969). Photograph courtesy of F. P. Gagliano.)

compete effectively with other welding technologies purely on an economic basis, at least in some cases.

D. Specific Examples of Laser Welding Capability

We illustrate the capabilities of laser welding by presenting more detail about two specific laser welding applications. The first example involves solution of an unusual problem, for which the very localized heating of the laser proved to be the crucial factor. The second example illustrates how laser welding can supplant a competing technology because it provides a cost advantage.

The first example is the sealing of a miniature detonator, with a laser weld made within 0.040 in. of a temperature-sensitive primary explosive. The hermetic sealing of miniature explosive devices such as detonators, small igniters, and actuators had been very difficult to achieve on a mass production basis with reliability.

The particular detonator was in the form of a cylindrical can with diameter 0.10 in. and length 0.25 in. The welding was accomplished using a repetitively pulsed CO_2 laser to weld a header to the can containing the high explosive. Both the can and header were 320 stainless steel. The laser beam was precisely controlled, yielding a spot diameter of 0.005 in. and a weld depth of 0.003 in. The heat was thus localized to a very small region. The weld was produced by overlapping spots with energy in the 0.1–0.3 J range and with duration around 40 μsec. The weld is shown in Figure 14-21, with the header at the top and the seam at the junction between the header and the can. The width of this seam is approximately 0.015 in. The marks from the overlapping pulses are visible. The weld was made in $\frac{1}{3}$ sec as the can was

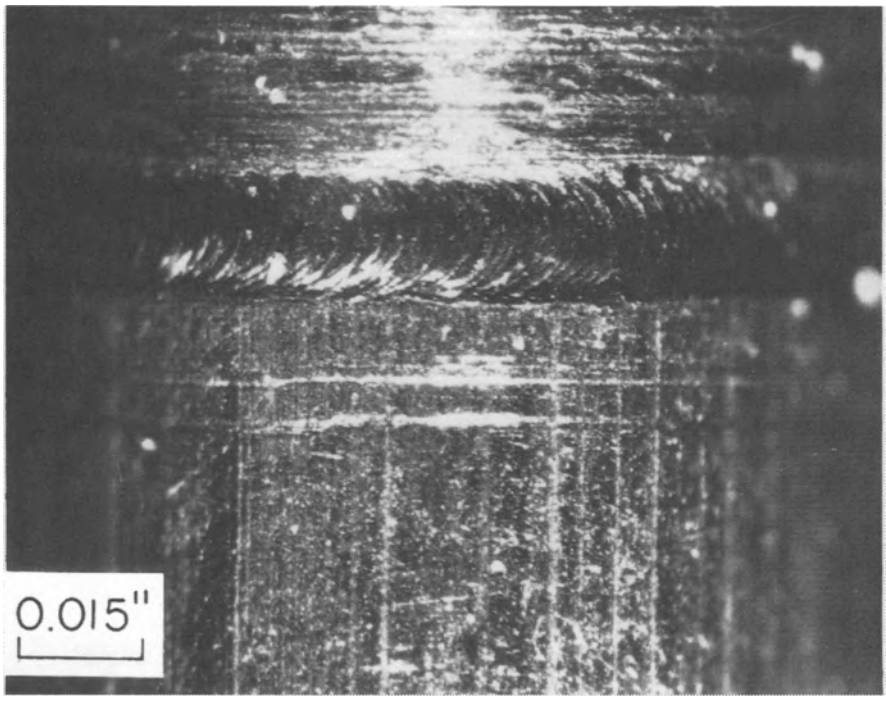


Figure 14-21 Seam weld around the periphery of a microdetonator can. The weld was made with a repetitively pulsed CO_2 laser. (From J. F. Ready, *Effects of High Power Laser Radiation*, Academic Press, New York, 1971.)

rotated under the focus of the beam, so that high-volume automated production was possible.

The welds are hermetic, and the device withstood adverse environments such as a 28-day temperature and humidity cycling test, 13,000 G shock in any direction, and thermal shock from –65 to 160°F. An important feature of the success of the laser welding was the very localized nature of the heating. Other methods of making the weld, including electron beam welding, were not successful because of the proximity of the weld to the temperature-sensitive explosive.

This example illustrates the use of laser welding in an application where it provided a technical advantage over other types of welding. The second example illustrates a case where another technique, electron beam welding, could also produce the welds effectively, but laser welding was selected because it offered higher throughput and lower unit cost. The example involves welding of small lithium batteries. The welding is performed by a 475-W repetitively pulsed CO_2 laser. The automated laser welder makes the final hermetic seal, welding the interface between the battery case and the terminal plate. The workstation, shown in Figure 14-22, accepts battery cells in magazines of 16. Several magazines are visible. A cell is

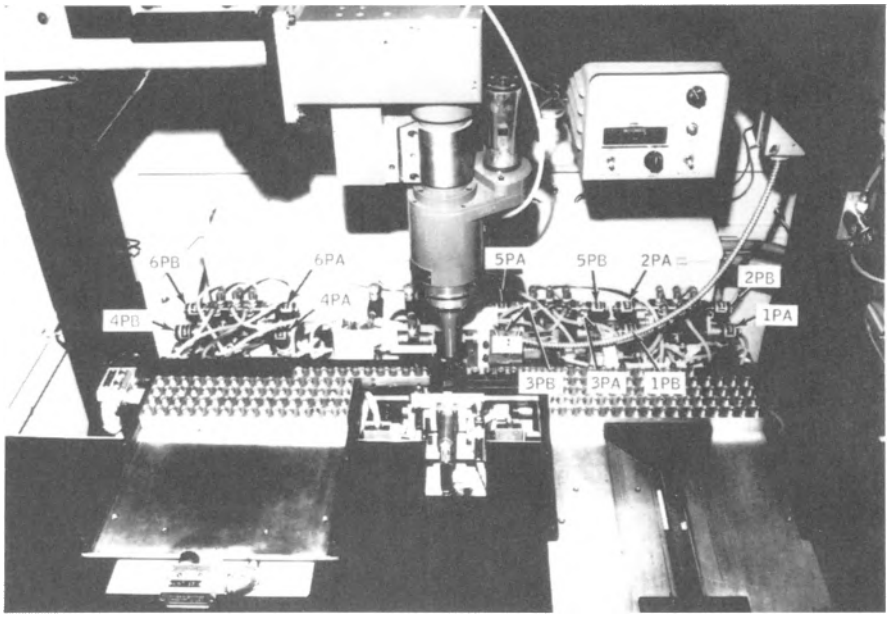


Figure 14-22 Automated workstation for laser welding of lithium batteries.

moved into weld position with linear indexing. The welding is done in an argon shield gas atmosphere. The cell is lifted from the magazine and electronically probed for presence and correct orientation. If these conditions are not met, the laser is not activated, and the next cell is brought into position. If the conditions are proper, the cell is clamped in a chuck and held motionless as the laser beam is rotated around the periphery of the top of the case. The laser beam is incident from the left through a metal pipe. It is turned 90° by a mirror and delivered through a lens, which is rotated to scan the beam around the top of the device. After being welded, the cell is lowered into the magazine, and the cycle is repeated. The actual weld is accomplished in 1.1 sec. The total cycle time is 4 sec, so that 900 batteries per hour are welded. These batteries also pass very stringent environmental and stress tests with a nearly 100 percent pass rate for the laser weld.

This welding operation had previously been performed with an electron beam. When a new production facility was needed to increase production rates, an economic analysis of competing welding technologies indicated that laser welding would provide the lowest unit cost.

E. Summary

There are now many practical applications in which laser welding has reached production status. Laser welding competes with many established techniques, like arc,

resistance, and electron beam welding. In many cases, laser processing offers some important advantage. Among the advantages of laser welding are the following:

- Laser welding is a high energy density process, so that the heating is very localized. This fact also influences several of the following characteristics.
- The heat-affected zone is very small. This can be crucial when a weld must be made near a heat-sensitive element, like a glass-to-metal seal.
- The total heat input into the sample is low. This can be an important factor for applications like welding electronic packages, or when low distortion of the part is desired.
- Because of the high energy density, welding rates can be high, so that the processing can be economical.
- No material contacts the workpiece, so that there is no contamination.
- Laser welding may be performed in the atmosphere or with a simple shielding gas; there is no need to move parts into and out of vacuum.
- Welding can be performed in otherwise inaccessible areas, for example, for repairs inside an enclosed vacuum tube.

These technical advantages, coupled with the fact that laser welding produces high-quality welds at high seam rates in a wide variety of metals, in a process that is easily automated and that achieves high throughput, have led to the adoption of laser welding for many practical industrial applications, especially in high-volume production.

The areas in which laser welding experiences some limitations include the relatively high cost of the equipment, the limitations on the thickness of the material that can be welded, and some difficulty in welding highly reflective metals. Table 14-4 compares some of the features of laser welding with competing techniques. The table includes some of the factors that might be included in making a choice of processing technology for a particular application.

Table 14-4 **Comparison of Characteristics of Welding Technologies**

	Welding technique			
Characteristic	**Laser**	**Electron beam**	**Resistance**	**Tungsten inert gas**
Weld quality	Excellent	Excellent	Good	Excellent
Heat-affected zone	Very small	Small	Medium	Large
Seam rate	High	High	Medium	Medium
Thickness	Limited	High	Medium	High
Equipment cost	High	High	Low	Low
Fixturing cost	Low	High	High	Medium

In recent years, laser welding in production applications has been expanding and displacing other techniques because its high throughput and simplified fixturing make laser processing cost-effective, even though laser welding equipment is relatively high in initial cost.

Currently, most laser welding applications are performed with either CO_2 lasers or Nd:YAG lasers, because these lasers are relatively mature and well developed. In the future, other developing types of lasers with high radiance and short wavelength may become practical welding sources. The development of lasers like the copper vapor laser, operating at a wavelength of 511 nm, with the capability of scaling to high average power, offers new possibilities. Because of the shorter wavelength, the beam may be focused to a smaller spot. This results in a smaller kerf, a higher irradiance at the workpiece, and attendant higher weld velocity. Figure 14-23 shows a comparison of the expected kerf and welding velocity for some different types of laser, both currently available lasers and projected developing types, such as large copper vapor and dye lasers. Because of the increased seam rates, the cost of the weld should also decrease. Currently, high-radiance laser welding is being investigated experimentally. It appears likely that in the future, high-radiance lasers may be used for industrial welding, in addition to the Nd:YAG and CO_2 lasers now widely used.

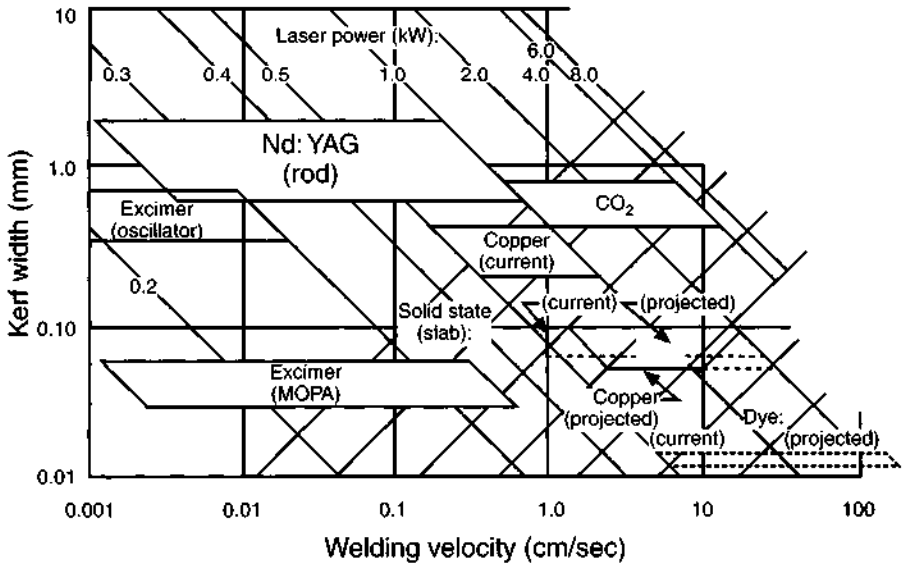


Figure 14-23 Comparison of kerf and welding velocity for current and projected laser types. (Figure courtesy of S. Hargrove, Lawrence Livermore National Laboratory.)

References

[1] Tables 14-1 and 14-2 are adapted from *The Industrial Laser Annual Handbook—1990 Edition*, D. Belforte and M. Levitt, eds. PennWell Books, Tulsa, OK, 1990.
[2] E. V. Locke and R. A. Hella, *IEEE J Quantum Electron.* **QE-10**, 179 (1974).
[3] C. L. M. Ireland, *Laser Focus/Electro-Optics*, p. 49 (November 1988).
[4] E. M. Breinan, C. M. Banas, and M. A. Greenfield, United Technologies Research Center Report R75-111087-3, 1975.
[5] F. P. Gagliano, Society of Manufacturing Engineers Technical Paper MR74-954, 1974.
[6] C. M. Banas, at the Society of Manufacturing Engineers Conference on Lasers in Modern Industry, Chicago, January 25–27, 1977.
[7] M. A. Saifi and S. J. Vahaviolos, *IEEE J. Quantum Electron.* **QE-12**, 129 (1976).
[8] Both examples here are from F. P. Gagliano, R. M. Lumley, and L. S. Watkins, *Proc. IEEE* **57**, 114 (1969).
[9] S. R. Bolin, *Proc. Tech. Program Electro-Optical Systems Design Conference*, Anaheim, CA, November 11–13, 1975.

Selected Additional References

A. Seam Welding: Subkilowatt Levels

C. M. Banas and R. Webb, Macro-Materials Processing, *Proc. IEEE* **70**, 556 (1982).
S. S. Charschan, ed., *Lasers in Industry*, Van Nostrand Reinhold, New York, 1972, Chapter 4.
W. W. Duley, *CO_2 Lasers: Effects and Applications*, Academic Press, New York, 1976, Chapter 6.
K. Hashimoto *et. al.*, Laser Welding Copper and Copper Alloys, *J. Laser Appl.* **3**, 21 (1991).
T. VanderWert, Low Power Laser Welding, in *The Industrial Laser Annual Handbook—1986 Edition* (D. Belforte and M. Levitt, eds.), PennWell Books, Tulsa, OK, 1986.

B. Welding with Multikilowatt Lasers

C. Banas, High Power Laser Welding, in *The Industrial Laser Annual Handbook—1986 Edition* (D. Belforte and M. Levitt, eds.), PennWell Books, Tulsa, OK, 1986.
A. S. Kaye *et al.*, Improved Welding Penetration of a 10-kW Industrial Laser, *Appl. Phys. Lett.* **43**, 412 (1983).
T. H. Kim, C. E. Albright, and S. Chiang, The Energy Transfer Efficiency in Laser Welding Process, *J. Laser Appl.* **2**, 23 (January/February 1990).
E. V. Locke, E. D. Hoag, and R. A. Hella, Deep Penetration Welding with High Power CO_2 Lasers, *IEEE J. Quantum Electron.* **QE-8**, 132 (1972).
L. Mannik and S. K. Brown, A Relationship between Laser Power, Penetration Depth and Welding Speed in the Laser Welding of Steels, *J. Laser Appl.* **2**, 22 (Summer/Fall 1990).
I. J. Spalding, Multikilowatt Laser Applications, in *The Industrial Laser Annual Handbook—1990 Edition* (D. Belforte and M. Levitt, eds.), PennWell Books, Tulsa, OK, 1990.

C. Spot Welding

H. N. Bransch *et al.*, Determining Weld Quality in Pulsed Nd:YAG Laser Spot Welds, *J. Laser Appl.* **3**, 25 (Fall 1991).
S. S. Charschan, ed., *Guide to Laser Materials Processing*, Laser Institute of America, Orlando, FL, 1993, Chapter 7.

Chapter 15 | Applications for Surface Treatment

Lasers have been used in a number of ways to modify the properties of surfaces, especially the surfaces of metals. Most often, the objective of the processing has been to harden the surface in order to provide increased wear resistance. In some cases, the goal has been to provide improved resistance to corrosion.

Laser applications in surface treatment have been dominated by the CO_2 laser, usually operating at multikilowatt levels, although there have been demonstrations of surface treatment using Nd:YAG, carbon monoxide, and excimer lasers.

We shall describe several different approaches to surface modification with lasers, including heat treating, glazing, surface alloying, and cladding.

A. Hardening

The most common method of surface treatment with lasers has been heat treating, in order to increase the surface hardness of the workpiece.

High-power lasers can be used for heat treating of metals. The laser used most often has been the CO_2 laser, because it provides the highest levels of continuous power. The laser beam irradiates the metal surface and causes very rapid heating of a thin layer of material near the surface. When the beam moves to a different area on the surface, the heat deposited in the thin layer will quickly be conducted away, and the heated area cools rapidly. This is essentially a quenching of the surface region. Certain metals, notably carbon steel, may undergo transformation hardening by heating followed by rapid cooling. This process yields an increase in the hardness of the surface.

The basic idea of laser hardening is to freeze a mixture of iron and carbon in a metastable structure called martensite, which is harder than the phases of iron–carbon that are normally present at room temperature. Figure 15-1 shows a so-called time–temperature–transformation (TTT) diagram for a carbon steel with 0.8 percent carbon. When the steel is heated above 723°C, the normal low-temperature phases

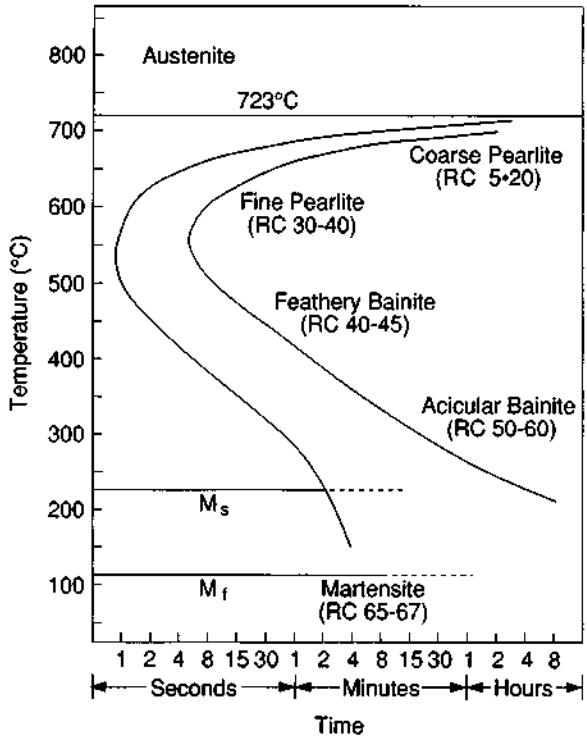


Figure 15-1 Time–temperature–transformation diagram for carbon steel with 0.8 percent carbon. RC denotes Rockwell C hardness. (From R. M. Brick and A. Phillips, *Structure and Properties of Alloys*, McGraw-Hill, New York, 1949.)

transform to a structure called austenite. Then as the steel cools, it transforms again to different phases, the nature of which depends on the cooling rate. Under slow cooling conditions, the material reassumes its usual low-temperature phases, which have Rockwell C hardness around 50. (The Rockwell C hardness test is a standard test based on the depth of indentation produced by a diamond indenter under a load of 150 kg.) The left curve in the TTT diagram indicates conditions at which the low-temperature phases begin to form. The right curve shows conditions for completion of the process. But if the cooling is rapid enough, that is, if the time–temperature profile passes to the left of the left curve, the normal low-temperature phases never begin to form. Instead, the harder martensite phase, with Rockwell C hardness of 65–67, begins to form at a temperature just over 200°C, at the line denoted M_s, and the transformation to martensite is completed at the line denoted M_f, just over 100°C. The martensite, which is a supersaturated solution of carbon in iron, has essentially been frozen in because of the rapid cooling.

In laser heat treating, the beam from a high-power laser is scanned over the metallic surface. The motion of the beam is too fast to allow melting to begin. The

surface layers are heated above 723°C, and austenite forms. When the beam moves on, the surface will cool extremely rapidly because of thermal conduction of the heat energy from the thin heated layer into the interior of the metal. This leads to transformation hardening of the areas that the laser beam has traversed. It is used to provide a hard surface in areas subject to wear.

As an example, Figure 15-2 shows some results for 1040 carbon steel hardened by a multikilowatt CO_2 laser. Part (a) shows the change in grain structure near the surface; part (b) shows the hardness as a function of depth. In this example, the CO_2 laser beam was scanned over the surface at a velocity of 10 ft/min, with a coverage of 75 in. ^{2}/min. The change in the material near the surface is apparent. The quenching has produced a finer grain size at the surface. The Rockwell C hardness of the base material was approximately 40, but in the 0.025-cm-thick layer near the surface, the hardness reached a value near 60. The rapid heating and quenching led to a significant increase in the hardness at the surface.

The hardened depth, often referred to as the case depth, depends on the thermal conductivity of the material. The surface temperature and the penetration depth may

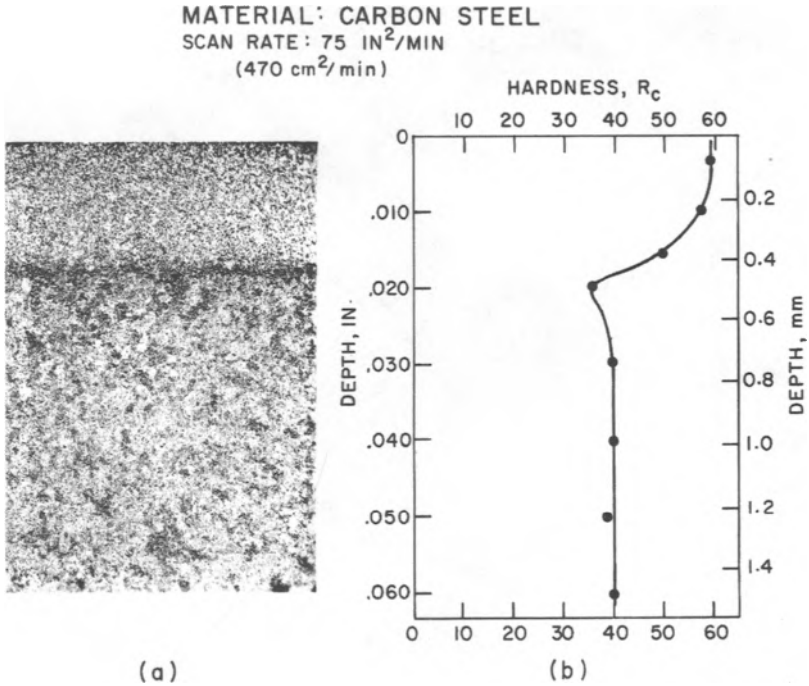


Figure 15-2 Results from heat treating of 1040 carbon steel by a multikilowatt CO_2 laser. (a) The grain structure as a function of depth, with the hardened surface layer apparent. (b) The Rockwell C hardness as a function of depth. (From E. V. Locke and R. A. Hella, *IEEE J. Quantum Electron.* **QE-10**, 179 (1974).)

be varied by adjusting the laser power, the focusing of the beam and the speed at which the laser beam is scanned across the surface. Greater hardening depths may be obtained by scanning at a slower rate, and shallower layers by scanning faster.

To cover a large surface area, one could use overlapping passes of the laser beam. But this may lead to annealing and a decrease of hardness in areas where the beam overlaps a previously hardened area. Various scanning procedures have been developed to overcome this problem and provide uniform hardening over a large area. One may traverse the beam relatively slowly in one direction while sweeping it back and forth rapidly in the transverse direction. One may obtain a uniform hardened depth when, at some depth in the metal, the oscillating pattern is transformed by thermal inertia into an essentially constant heat flow into the material.

Some results obtained with a 1500-W CO_2 laser beam that oscillated transversely to the scan direction are shown in Figure 15-3. This figure presents the trade-off between coverage rate and hardened depth in gray cast iron for two different rates of oscillation as indicated. In the top right portion of the figure, the movement of the beam is slow enough that surface melting occurs. Thus, oscillation frequency can influence whether melting occurs or not.

Another approach to coverage of a large area is to use a uniform beam profile to cover the width of the workpiece in one pass. This is the approach that is most commonly used. Because laser beams typically have round profiles with a variation of the power across the beam, heat treating applications frequently use beam homogenizers (also called beam integrators) to achieve uniform irradiance on the surface.

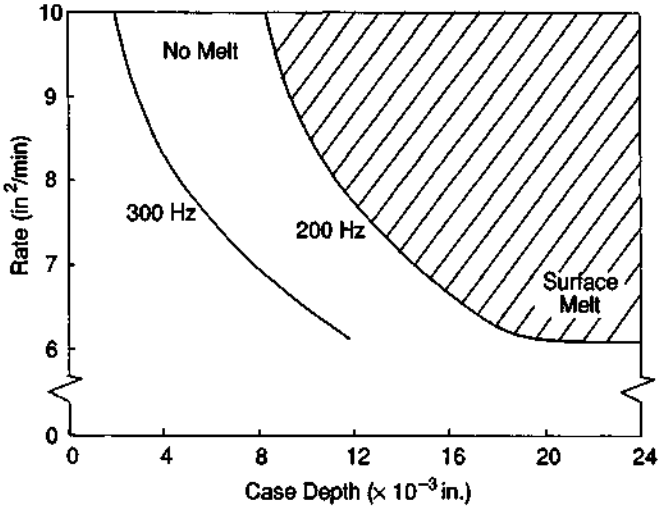


Figure 15-3 Rate of surface coverage as a function of case depth for laser hardening of cast iron by a CO_2 laser beam with power of 1500 W focused to a spot size of 0.05 in. The curves show coverage rates at the indicated frequencies of transverse oscillation of the beam. (From S. L. Engel, *American Machinist*, p. 107 (May 1976).)

The beam homogenizer, mentioned in Chapter 5, is a multifaceted mirror, with square facets. The different facets produce reflections that are directed at different angles so as to form a square beam with relatively constant irradiance across the beam profile. Such a beam is well suited for applications like heat treating.

The results of laser heat treating are frequently presented in a format called a Harth diagram. This presents curves of the laser beam power and scan speed necessary to obtain a constant hardened depth. An example is shown in Figure 15-4, which shows the relation between laser power and scan speed to obtain a hardened depth of 0.056 cm in carbon steel. The beam diameter is the parameter for the different curves. For a different hardened depth, there would be a different set of curves. The Harth diagram allows one to determine the parameters required to obtain a specified result.

Figure 15-5 shows how the coverage rate varies with hardened depth for hardening of gray cast iron with a 2.5-kW CO_2 laser beam. Results such as these allow one to trade off coverage rate for increased penetration depth.

All metals that can be transformation hardened can be hardened with a laser. Examples include carbon steels, tool steels, and cast iron. The austenite to martensite transition discussed earlier is the most common type of laser hardening.

One desires to avoid surface melting, so that the laser irradiance at the surface is much lower than what is used for welding and cutting. An irradiance around

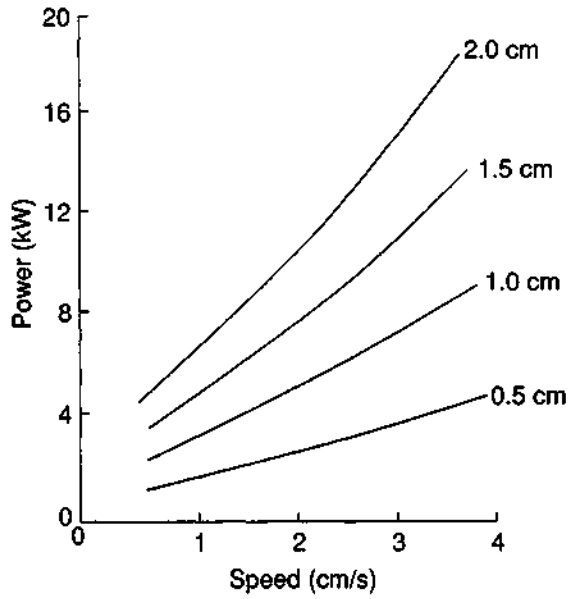


Figure 15-4 Laser power versus scan speed required to produce a hardened depth of 0.056 cm in carbon steel. The different curves were obtained at the indicated values of beam diameter. (Data of V. Gregson.)

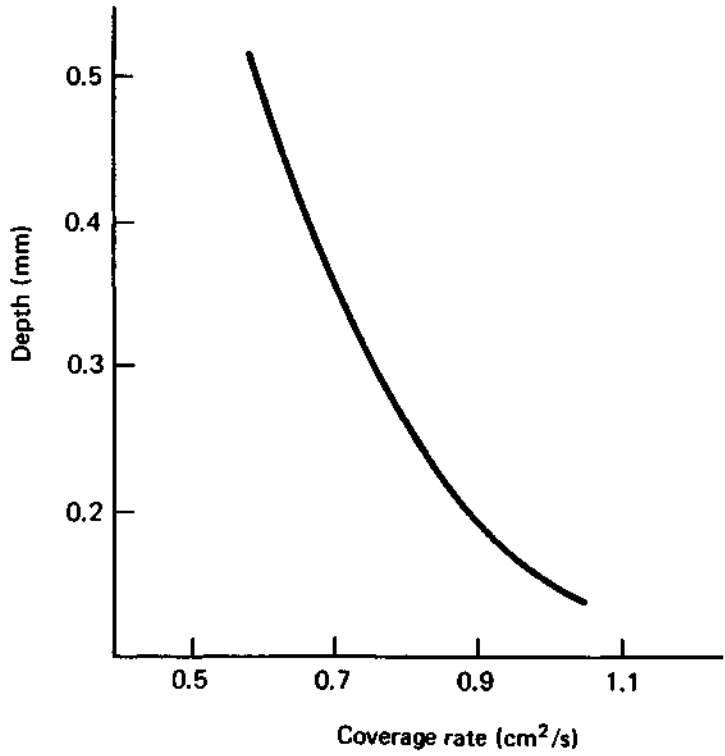


Figure 15-5 Hardened depth as a function of coverage rate for hardening of gray cast iron with a 2.5-kW laser beam focused to a diameter of 5 mm. (From J. T. Luxon and D. E. Parker, *Industrial Lasers and Their Applications*, Prentice Hall, Englewood Cliffs, NJ, 1992.)

10^4 W/cm^2 is typical. Practical hardening depths are perhaps around 0.03 in., with coverage rates of a few tens of square inches per minute at laser power levels of a few kilowatts. These rates are high enough to make laser hardening economically feasible.

The laser energy must be absorbed well at the surface. Because the CO_2 laser is used for almost all laser hardening, the high reflectivity of metal surfaces in the infrared can be a problem. The absorption of the beam can be increased by a thin layer of highly absorbing material on the surface. The coating must be very thin, perhaps around 0.001 in., to allow effective transfer of the energy to the workpiece. If the coating is too thick, penetration of the heat is reduced, and the case depth is lessened.

A variety of coating materials have been investigated and found to provide very high values of absorption. The materials include polycrystalline tungsten, cupric oxide, graphite, molybdenum disulfide, and manganese phosphate. Of these, manganese phosphate may be the one most easily applied in an industrial plant, although it does not have quite as high absorption as the other materials. In some cases, use

of high-temperature automobile manifold paint has proved satisfactory and easy to apply.

Thus, a number of coating materials are available to increase the absorption of the surface to the 80–90 percent range at the CO_2 laser wavelength. This enables one to obtain good results in surface hardening of transformable metals.

An example of a hardened band produced on the surface of a gray cast iron sample is shown in Figure 15-6. The hardening was produced by a 1000-W CO_2 laser. The beam was scanned across the surface at 10.2 mm/sec and was rapidly oscillated by a mirror in the transverse direction to widen the hardened area and to produce more uniform hardening. The width of the hardened band is 6.8 mm, and the hardened depth is 0.53 mm. The hardened material consists of a matrix of fine martensite containing flake graphite. The hardness in the hardened zone is 58–62 (Rockwell C).

Laser hardening is suitable for applications that require high hardness with relatively shallow case depth on selected surface areas, with no dimensional distortion. It can be used for geometries with irregular shapes, grooves, lines, and gear teeth. The automotive industry, in which wear resistance on selected surfaces is desired, has been the most important application area for laser hardening. Laser hardening of steering gear housings, crankshafts, turbocharger rotor shafts, cylinder blocks, and diesel engine cylinder liners has reached routine production status, and hardening of many other automotive components has been investigated. But laser hardening has not been adopted for practical production use to the same extent that laser welding and laser cutting have been. Only a relatively small fraction of lasers used for industrial processing are used for hardening.

The most common competing technique for surface hardening of metals is heating in an induction furnace. Induction heating can produce hardened layers with thickness up to a few millimeters. For applications in which the depth of hardening does not have to be great and where the part geometry is uneven, so that induction processing is not very suitable, laser hardening provides a good alternative. Laser hardening can provide more selective hardening than induction heating, and it yields somewhat higher values of hardness because of the more rapid quenching. Laser hardening also produces a substantial increase in surface hardness at a reasonably high rate with minimum distortion of the part. However, equipment for laser

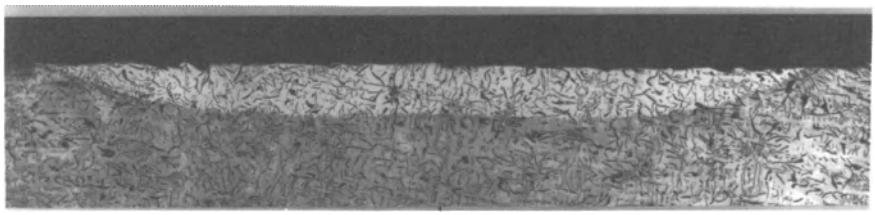


Figure 15-6 Hardened zone produced on the surface of gray cast iron by CO_2 laser transformation hardening. (Photograph courtesy of M. J. Yessik.)

hardening is more expensive, especially because it usually requires a multikilowatt laser. Thus, it has not replaced induction hardening except for selected applications.

We turn now to several other methods for providing hardened surfaces through laser processing, including glazing, alloying, and cladding. These methods are all less well developed than transformation hardening, but they appear to offer promise for future exploitation.

B. Glazing

Laser glazing [1] involves some surface melting. As the beam from a multikilowatt CO_2 laser is scanned over a surface, a thin melt layer is produced under proper conditions of irradiance and traverse speed. The interior of the workpiece remains cold. After the beam moves on, resolidification occurs very rapidly. The surface layer is quickly quenched. As a result of this process, one may produce surface microstructures with unusual and possibly useful characteristics. The grain size near the surface is very small, because of the high quench rate. The surface structure can appear glassy; hence, the name laser glazing. This technique is applicable both to metals and ceramics. It appears to be controllable and reproducible.

Laser glazing can produce surfaces that are amorphous or that have a glassy, noncrystalline structure. Such surfaces can have increased resistance to corrosion.

One example of laser glazing is surface melting of cast iron with rapid resolidification. This produces a thin surface layer of very hard material called white cast iron, which can provide excellent wear resistance. Another example that has been demonstrated is the glazing of aluminum bronze, which leads to surface structures with enhanced corrosion resistance.

C. Laser Alloying

Surface alloying involves spreading of a powder containing alloying elements over the surface to be hardened. Then, the laser beam traverses the surface. The powder and a thin surface layer melt and intermix. After passage of the beam, the surface resolidifies rapidly. A thin layer containing the alloying elements remains, with hardness greater than that of the original material.

An alternative method of supplying the alloying elements uses a gas jet to blow the powder onto the melted surface.

The laser typically is a multikilowatt CO_2 laser in order to achieve coverage of large areas. The treating rate is somewhat slower than for transformation hardening, because some surface melting is required.

Surface alloying can produce surfaces with desirable properties on relatively low cost substrate materials. For example, one could produce a stainless steel surface on low carbon steel by alloying nickel and chromium.

Alloying requires a high value of heat input, controlled to produce the desired depth of surface melting. A laser is the only practical heat source for surface alloying.

In one study [2], the elements carbon, boron, chromium, and silicon were mixed in powder form on the surface of a valve seat. The powder was melted and alloyed into the surface by an 18-kW CO_2 laser beam. The treating and surface hardening took less than 1.5 sec for a valve seat.

Despite considerable promise and research interest, laser alloying has not made the transition to widespread production use.

D. Laser Cladding

Laser cladding involves bonding of an overlay material to the surface to be hardened. The overlay material may be pre-positioned or it can be continuously fed. The laser beam melts a thin surface layer, similarly to the case of laser glazing. The overlay material is then metallurgically bonded to the surface. In contrast to glazing, the overlay material does not intermix with the surface. Only enough of the surface melts to form the bond with the overlay material. The process is also called hardfacing. Again, the laser would probably be a multikilowatt CO_2 laser.

Laser cladding usually involves covering a relatively inexpensive substrate material with a more expensive alloy that will increase the resistance of the part to wear or corrosion. Cladding allows the bulk of the part to be made with inexpensive material, while providing the surface with desirable properties associated with the more expensive cladding material. Laser cladding has resulted in surfaces with very good finish, good homogeneity, and very low porosity. Common materials used for cladding include carbides, iron-based alloys, nickel base alloys, and cobalt base alloys.

Conventional techniques for cladding have employed tungsten-inert gas welding, plasma spraying, and flame spraying. These approaches have encountered problems, including porosity of the cladding, uneven cladding thickness, dilution of the cladding alloy because of the large heat input required, and distortion of the workpiece by the high heat input. Laser cladding has reduced these problems and produced cladding layers with good uniformity, good dimensional control, and minimal dilution of the cladding alloy.

In one example [3], carbon steel was hardfaced with a cobalt alloy by a 12-kW CO_2 laser beam. The alloy was overlaid on the steel as a powder layer 2.5 mm thick. After melting and bonding by the laser, the alloy layer was 1.5 mm thick and had very little dilution from the steel. The resulting surface had excellent wear resistance.

Laser cladding has been applied to hardening of steam generator turbine blades and reached pilot production status [4]. The blades had a machined pocket, into which powdered cobalt-based metal (Stellite 6) was fed ahead of the laser beam. The resulting cladding layer was well bonded, and the blades exhibited improved fatigue characteristics compared with blades formed by the conventional silver brazing process.

Laser cladding also has been investigated for increasing the oxidation resistance of turbine blades [5]. The nickel-based substrate (Rene 80) was clad with a coating

containing nickel, aluminum, chromium, and the rare earth element hafnium. The powder was fed by gravity to the interaction region, where the cladding was performed by the beam from a 10-kW CO_2 laser. The clad substrates showed improved oxidation characteristics compared with the original alloy.

As is the case with laser glazing, there have been numerous successful demonstrations of laser cladding, but relatively few production uses.

E. Specific Examples of Laser Heat Treating Capability

In one of the first production applications of laser transformation hardening, dating to 1978, locomotive diesel cylinder liners have been hardened with a 5-kW CO_2 laser beam. Laser heat treating was chosen because it offered the advantages of minimal distortion of the bore of the liner and localized heating that did not affect nearby brazed joints. The laser produces a localized hardening in a spiral track. The hardened depth has been great enough to withstand normal service plus reconditioning. The presence of the hardened track reduces scuffing of the liner by the piston and, as a result, extends engine life by 30 percent. The multistation processing facility has a capacity of 120 units per shift, and it has been operated three shifts per day.

Perhaps the best-known application of laser transformation hardening has involved hardening of cast iron automatic steering assemblies with CO_2 laser radiation [6]. The laser hardening process was adopted because it could meet the engineering requirements with less hardened surface than conventional processing. The required wear resistance was provided by four and one-half tracks of laser hardening around the interior diameter of the assembly. Because of the minimal surface area involved, distortion of the part was reduced and post-treatment machining was eliminated. Another desirable characteristic was the precise placement of the wear resistant areas. At its peak utilization, 22 laser hardening stations were in use for this application.

In another significant production application, laser hardening is used to enhance the lifetime of blades for steam turbines [7]. These blades are eroded by the impact of water droplets during turbine operation. Other methods of hardening the blades, such as flame hardening or cladding with a hard metal like stellite, all have significant disadvantages.

Laser hardening has successfully been applied to turbine blades and has been in production in this area for some time. The blades are hardened with a 6-kW CO_2 laser. The hardening speed is between 150 and 400 mm/min, with a width of 30 mm, so that the coverage can be as high as 120 cm^2/min. Testing of the hardened blades indicated that the mass loss after 1.5 million revolutions was very substantially reduced compared with that for other hardening technologies. Laser hardening provided excellent wear properties and high fatigue strength. Extended life tests in steam turbines showed that the laser-hardened blades meet all requirements. The laser-hardened blades have been in use in steam turbines in many countries for a number of years.

References

[1] E. M. Breinan, B. H. Kear, and C. M. Banas, *Phys. Today*, p. 44 (November 1976).
[2] R. Hella, at the 1975 IEEE/OSA Conference on Laser Engineering and Applications, Washington, DC, May 28–30, 1975.
[3] G. J. Bruck, at the American Welding Society Conference, New Orleans, LA, April 1988.
[4] G. J. Bruck, in *Laser Beam Surface Treating and Coating* (G. Sepold, ed.), *SPIE Proc.*, Vol. 957 (1988), p. 14.
[5] S. Sircar, C. Ribaudo, and J. Mazumder, in *Laser Beam Surface Treating and Coating* (G. Sepold, ed.), *SPIE Proc.*, Vol. 957 (1988), p. 29.
[6] D. Belforte, *Industrial Laser Review*, p. 4, (December 1988).
[7] E. Brenner and W. Reitzenstein, *Industrial Laser Review*, p. 17 (April 1996).

Selected Additional References

S. S. Charschan, ed., *Guide to Laser Materials Processing*, Laser Institute of America, Orlando, FL, 1993, Chapter 6.
J. Mazumder, Laser Heat Treatment: The State of the Art, *J. Metals*, p. 18 (May 1983).
F. D. Seaman, Laser Heat Treating, in *Industrial Laser Annual Handbook—1986 Edition* (D. Belforte and M. Levitt, eds.), PennWell Books, Tulsa, OK, 1986.
F. D. Seaman, New Developments in Laser Surface Modification, in *Industrial Laser Annual Handbook—1990 Edition* (D. Belforte and M. Levitt, eds.), PennWell Books, Tulsa, OK, 1990.
E. N. Sobol, *Phase Transformations and Ablation in Laser-Treated Solids*, Wiley, New York, 1995, Chapter 4.

Chapter 16 | Applications for Material Removal: Drilling, Cutting, Marking

In the last two chapters, we described applications for laser-induced heating and melting. In those applications one generally tries to avoid vaporization and material removal. In this chapter, we emphasize applications that require material removal. Generally, this means operation at higher levels of laser irradiance delivered to the workpiece compared with the operating levels for welding and surface treatment. We will describe a variety of different applications, including hole drilling, cutting, and marking, but we shall not discuss the uses in fabrication of electronic components. This will be left to the next chapter.

A. Laser-Induced Material Removal

We have already discussed some of the considerations involving laser interactions and material removal in Chapter 12. In this chapter we shall emphasize aspects of material removal as they relate to applications.

There are several different types of lasers of interest for material removal. The list of lasers is different from those used for welding, because one may use pulsed lasers with higher levels of power. The useful range of pulse duration is shorter than that required for welding, but the irradiance must be somewhat higher. Thus, CO_2 TEA lasers, Q-switched Nd:YAG lasers, and excimer lasers can be useful for material removal, but they are not often used for welding. Continuous lasers are less often used for applications like hole drilling or material removal than for welding and heat treating. Table 16-1 lists some of the lasers often used for material removal applications. Of the lasers in the table, most material removal applications have been done with Nd:YAG or CO_2 lasers. Excimer lasers are now used widely because of the high absorption of many materials in the ultraviolet region. Ruby and Nd:glass lasers were used in many early studies, but they are less common now because of their low pulse repetition rates.

Table 16-1 **Lasers for Material Removal**

Laser	Wavelength (μm)	Typical applications
Nd:YAG (pulsed)	1.06 and 0.532	Hole drilling, marking
Nd:YAG (Q-switched)	1.06 and 0.532	Trimming, thin film removal
CO_2 (pulsed)	10.6	Cutting, hole drilling
CO_2 (TEA)	10.6	Balancing
Excimer	0.19–0.35	Marking, semiconductor processing
Ruby	0.6943	Drilling
Nd:glass	1.06	Drilling

In applications involving material removal, the properties of the workpiece are important, especially the thermal diffusivity, the latent heat of vaporization, and the surface reflectivity. Under conditions where the material is vaporized, the amount of material that can be removed is limited primarily by the latent heat of vaporization. If we assume that all the incident laser energy is used to raise the material to its vaporization temperature and to supply the latent heat of vaporization, we may derive a simple estimate for the maximum amount of material that may be vaporized. The procedure simply equates the energy supplied by the laser to the energy needed to heat a given amount of material to its vaporization temperature and to vaporize it. This is sometimes called an energy balance model of the process. Using this model, we obtain the maximum depth D of a hole that may be drilled by a laser pulse of energy E_0 as

$$D = \frac{E_0}{A\rho[c(T_B - T_0) + L]} \tag{16.1}$$

where c is the heat capacity per unit volume, T_B the vaporization temperature, T_0 the ambient temperature, L the latent heat of vaporization per unit mass, ρ the density, and A the area to which the beam is delivered. This equation neglects the latent heat of fusion, which is much smaller than the latent heat of vaporization. If, as an example, we apply this equation to aluminum, the relevant values are

$$\rho = 2.7 \text{ gm/cm}^3, \qquad c = 0.97 \text{ J/gm-deg}$$

$$T_B - T_0 = 2447°\text{C}, \qquad L = 10{,}900 \text{ J/gm}$$

Then a laser pulse with an energy of 10 J, focused to a spot with an area of 10^{-3} cm^2, will produce a hole with a maximum depth of 0.28 cm, according to Equation (16.1). In practice, the depth of a laser-drilled hole is influenced by other factors. Loss of laser energy by surface reflection and by thermal conduction into the interior will decrease the hole depth. Flushing of molten material, which is not completely vaporized, will increase the hole depth. The energy balance model will indicate general capabilities and limitations, often to within a factor of two. It thus may be used as a quick estimate for feasibility of a specific application with a particular laser.

Data on the amount of material removed by Nd:glass laser pulses from aluminum and iron are shown in Figure 16-1. These curves will give an estimate of how much material may be removed from representative materials. They also show an interplay between latent heat and thermal diffusivity. The curves for the two metals cross. At relatively low fluence, loss of heat into the interior of the workpiece is important. The thermal diffusivity of aluminum is higher than that of iron, so that the loss is more severe for aluminum and less aluminum is vaporized. As laser energy increases, the material reaches vaporization conditions more rapidly and there is less time for thermal conduction. The loss due to thermal conduction becomes less important than the latent heat of vaporization. Because iron has higher latent heat of vaporization than aluminum, the holes in aluminum are deeper than for iron near the right edge of the figure.

The dominant process in material removal by lasers may often be simple vaporization, with absorption of the laser energy at a continually retreating surface. The vaporized material can simply diffuse away into the surrounding atmosphere, without further interaction with the beam. The upper limit for material removal in this fashion is given by the energy balance model, as expressed by Equation (16.1). Thermal conduction can remove heat and thus reduce the amount of material removed. The data in Figure 16-1 were obtained under conditions where thermal

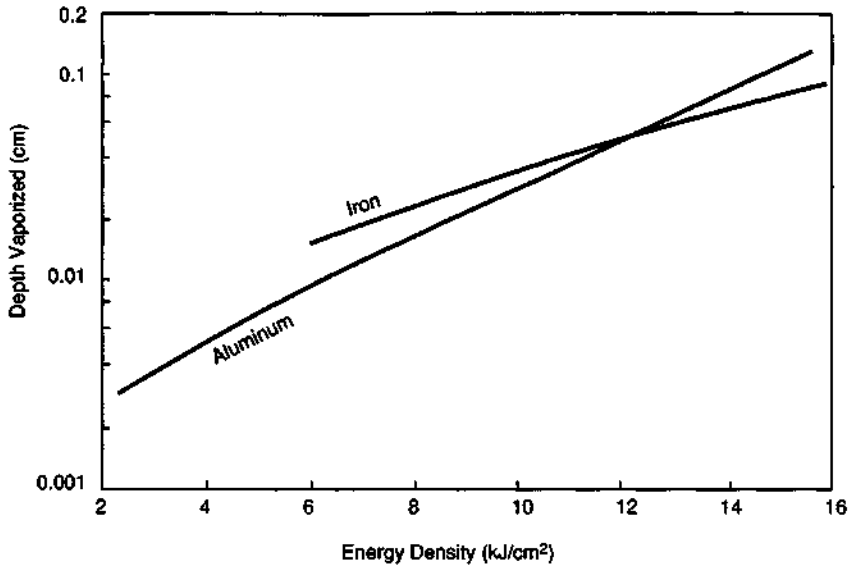


Figure 16-1 Experimental values of depth vaporized by a 700 μsec pulse from a Nd:glass laser as a function of laser energy per unit area incident on iron and aluminum targets. (These data are adapted from measurements of mass removal by V. B. Braginskii, I. I. Minakova, and V. N. Rudenko, *Sov. Phys. Tech. Phys.* **12**, 753 (1967).)

conduction was significant. The values given in this figure are smaller than what would be calculated with Equation (16.1).

Other mechanisms can also be operative, as Chapter 12 described. Mechanisms that can increase the amount of material removal are flushing of liquid material [1] and particulate emission because of subsurface explosions [2]. Absorption of laser energy in the blowoff material can shield the surface and reduce the amount of material removal. Thus, if one is investigating an application involving laser material removal, an experimental study of the process must be conducted to determine what processes are important and to learn the optimum regime of laser parameters to use.

Considerations about surface reflectivity are similar to those for welding. Laser energy may be lost by reflection from the surface. In the far infrared, where metallic surface reflectivity is high, it becomes more difficult to achieve effective vaporization. CO_2 lasers can accomplish vaporization with reasonable efficiency because the surface reflectivity does drop after the surface disruption begins. The energy then is coupled into the workpiece reasonably well. This phenomenon has already been described in Chapter 12.

For material removal from nonmetallic materials, the absorption at the laser wavelength is important. Materials like glass, quartz, and many plastics are transparent in the visible and near infrared regions, and so a Nd:YAG laser will not deposit much energy. But they are highly absorbing in the far infrared region, so that a CO_2 laser will deposit most of its energy. Such behavior is also typical of many organic materials, most of which have high absorption in the far infrared region. Ceramics are also generally highly absorbing at the CO_2 laser wavelength. Silicon has higher absorption in the near infrared region, so that a Nd:YAG laser will deposit more energy in silicon than will a CO_2 laser. At ultraviolet wavelengths, many nonmetallic materials have extremely high absorption. This is a factor that favors use of excimer lasers, which can deliver a very high value of energy per unit volume in the region near the surface of a workpiece. Thus, when one chooses a laser for material removal, it is important to consider the absorption characteristics of the material as a function of wavelength.

In the next sections, we describe some of the most important applications for laser-based material removal, including drilling, cutting, and marking.

B. Hole Drilling

Hole drilling has developed as one of the most important laser applications in material processing. Laser drilling of holes in ceramic substrates has become widely used. Hole drilling in metals is useful in a variety of areas, such as production of tiny orifices for nozzles and controlled leaks, apertures for electron beam instruments, cooling holes in aircraft turbine blades, and pinholes for optical applications. Production of tiny orifices, with dimensions as small as 0.00015 in., through thin metallic sheet has been demonstrated, although for drilling thicker stock, a minimum hole diameter around 0.001 in. is common. Precise, close patterns can be accomplished

with no heating of adjacent areas and no contamination. The heat-affected zone around the holes is very small, typically 0.001–0.002 in. Very small customized holes can be made in samples that conventional methods cannot handle well.

Hole drilling typically employs pulses with shorter duration than those for welding. Pulse lengths in the region of several hundred microseconds, or even less, are generally used. Rarely would a pulse longer than 1 msec be used for drilling. Also, submicrosecond duration pulses are not often used, because of generation of an absorbing plume if the irradiance becomes too high. Table 16-2 shows experimental values for the depth of laser-produced holes in several metals for two cases: a submicrosecond high-power pulse, and a longer, lower-power pulse. The results were obtained with the same Nd:glass laser, with and without Q-switching. The Q-switched pulse removed less material. Even noting that the shorter pulse contained less energy, it is apparent that the short pulses were less effective in material removal. The amount of material removed per unit incident energy was smaller for the Q-switched pulses.

The lasers most commonly used for hole drilling are the CO_2 laser and the Nd:YAG laser. Because of their shorter wavelength, Nd:YAG lasers are suited to drilling smaller holes. For typical thicknesses of metal, the range of hole diameters possible with a Nd:YAG laser lies in the range 0.001–0.060 in., whereas for CO_2 lasers, the range is 0.005–0.050 in. For precise applications in very thin stock (0.005 in. thick), holes as small as 0.00015 in. diameter have been drilled with Nd:YAG lasers.

Very large holes may be formed by trepanning, that is, by drilling a series of overlapping holes around the circumference of a circle so that the central slug of material drops out. For most practical hole drilling, the upper limit of thickness that can be drilled is in the range 0.5–1.0 in.

The use of shorter-wavelength lasers allows drilling of holes with smaller diameter and larger aspect ratio (ratio of depth to diameter). In one example [3], a copper vapor laser operating at 511 nm drilled holes nearly 1 mm deep with a diameter around 20 μm in a copper–zinc alloy, yielding an aspect ratio around 50.

Table 16-2 **Depths of Laser-Produced Holes**[a]

Material	Depth for 10^9 W/cm^2, 44 nsec pulse (cm)	Depth for 5000 J/cm^2, 600 μsec pulse (cm)
Stainless steel	0.00011	0.061
Brass	0.00025	0.078
Aluminum	0.00036	0.078
Copper	0.00022	0.090
Nickel	0.00012	0.058

[a]From J. F. Ready, *J. Appl. Phys.* **36**, 462 (1965).

Laser-drilled holes have some taper, as Figure 16-2 illustrates. This figure shows the cross sections of holes vaporized in aluminum by ruby laser pulses. Each cross section represents a new hole drilled in a fresh area. As the energy increases, both hole depth and diameter increase. The figure shows the typical shape of a laser-drilled hole, with a definite taper. The diameter is greatest at the entrance and decreases deeper in the hole. At the highest energy, the exit of the hole is enlarged again, probably as a result of the pressure exerted by vaporized material in the hole. The presence of taper in laser-drilled holes is a persistent feature. It may be minimized by use of a beam with a good spatial profile, but it will not be completely eliminated.

One may employ repeated pulses delivered to the same spot to deepen a hole. Figure 16-3 shows the relation between the number of pulses from a ruby laser and the depth of the hole drilled in a sapphire target. The depth increased until about 30 pulses, after which it did not increase further. Hole drilling in sapphire is of interest because it represents a material that is very hard and difficult to drill by nonlaser techniques.

Figure 16-4 shows the maximum hole depth that could be produced by a Nd:YAG laser in high nickel alloys, as a function of hole diameter. The figure summarizes the results of a parametric study of drilling in which several relevant factors, including

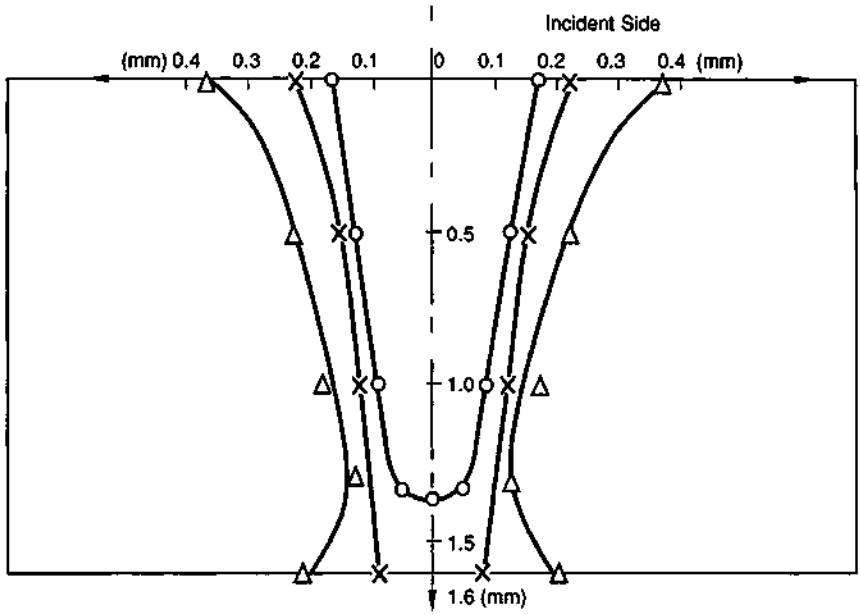


Figure 16-2 Change in the cross section of holes drilled by a ruby laser in 1.6-mm-thick aluminum plate as laser energy increases. The normal pulse ruby laser beam was focused with a 30 mm focal length lens. The points represent the following laser energies: ○, 0.36 J; ×, 1.31 J; △, 4.25 J. (From T. Kato and T. Yamaguchi, *NEC Res. Dev.* **12**, p. 57 (October 1968).)

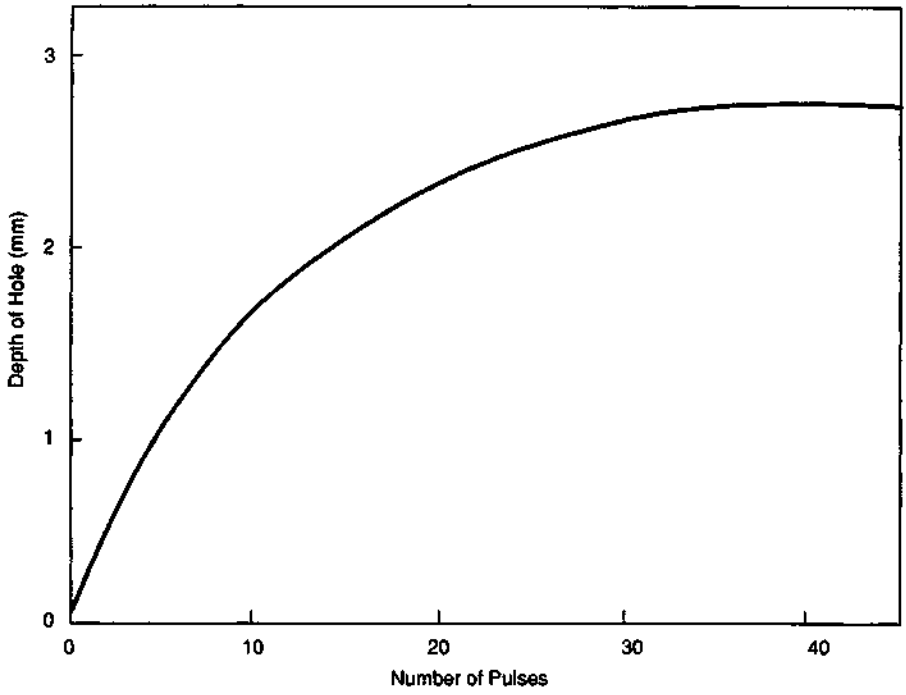


Figure 16-3 Depth of hole drilled in sapphire as a function of the number of pulses delivered to the same spot. The hole was drilled by ruby laser pulses with 2 J energy and 1.5 msec duration, focused by a 3 cm focal length lens. (Data from S. Shimakawa, Japan Society for Laser Material Processing.)

lens focal length, depth of focus, laser energy, and the number of pulses, were varied. The results are applicable to drilling of holes in air-cooled aircraft engine components, including turbine blades and vanes, nozzles, shrouds, and combustors. The results obtained with laser drilling compare favorably with the capabilities of other possible drilling processes. Holes with small diameter (< 0.01 in.) and large aspect ratio (ratio of depth to diameter) may be drilled. According to the figure, at a diameter of 0.01 in., the aspect ratio may exceed 60.

Practical hole drilling with lasers does suffer from some limitations. One limitation is associated with the limited depth of penetration, which arises from the limited amount of laser energy available in a short pulse. If one attempts hole drilling with longer pulses or with a continuous laser, the heat is conducted over a larger volume of material, and much of the advantage of using lasers is lost. Metallic plates a few millimeters thick may easily be pierced, either by a single pulse or by a series of pulses with lower energy. The maximum depth for which such piercing is practical, while maintaining a small diameter, is in the approximate range 12–25 mm.

A second limitation is the quality of the hole. An example of the cross section of a typical hole in a massive sample appears in Figure 16-5. This hole, drilled in brass

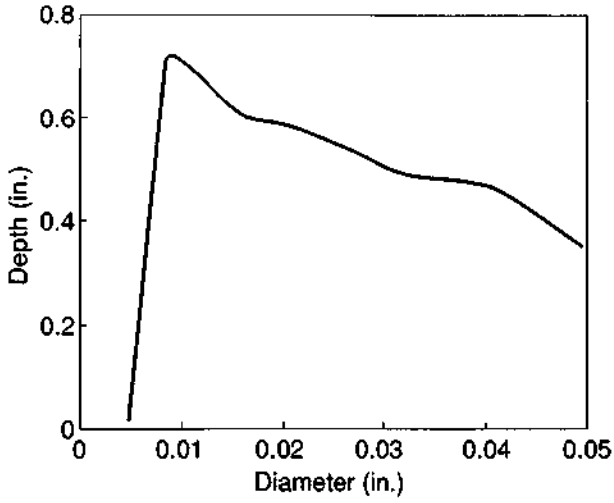


Figure 16-4 Capability for hole drilling in high nickel alloys for aircraft engine applications. The curve shows the largest depth that could be drilled at a specified diameter. (From A. U. Jollis, Small Hole Drilling and Inspection with Pulsed Laser Systems, *AIAA/SAE/ASME Fifteenth Joint Propulsion Conference*, June 18–20, 1979, Las Vegas, NV.)

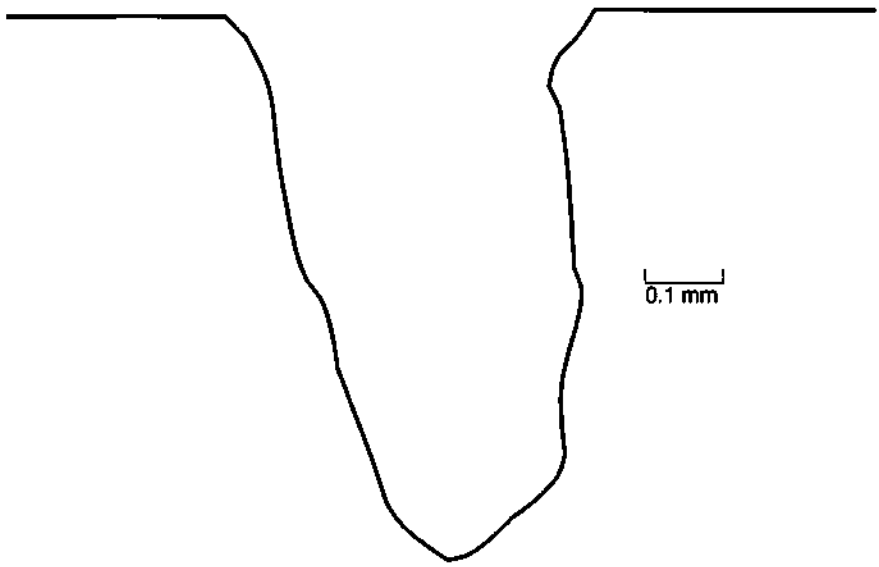


Figure 16-5 Cross section of hole produced in massive brass by a single 5 J pulse from a Nd:glass laser focused on the target surface by a 31 mm focal length lens. (From J. F. Ready, *Effects of High-Power Laser Radiation*, Academic Press, New York, 1971.)

by a single 5 J pulse from a Nd:glass laser, shows a number of defects, including rough sides, lack of roundness, offset of the center line of the hole from the normal to the surface, and a general taper, as mentioned before. These defects are difficult to eliminate completely and may disqualify the laser from some applications in which perfection of the hole is important.

Efforts to reduce taper have emphasized careful control of the spatial and temporal profile of the beam. The laser should operate in a good TEM_{00} mode, with no "hot spots." In this way, some improvement in the shape of laser-drilled holes is possible, but one apparently cannot eliminate the taper completely.

To reduce roughness of the sides of a hole drilled through a plate, one may blow gas through the hole as it is being formed. In hard, brittle materials, like diamond, a final finishing step may be used to smooth the sides of the hole.

Another common limitation involves the recondensation of material around the entrance of the hole, forming a crater-like lip for the hole. This is shown in Figure 16-6. As the figure shows, the exit side is clean, but the entrances are ringed with recondensed material. These lips are formed when material vaporized and ejected from the hole recondenses on the first cool surface that it strikes. One may generally remove this material mechanically or, alternatively, apply a coating to the sample that may easily be stripped after the drilling, taking the lip with it. But these alternatives involve additional steps and expense to the process.

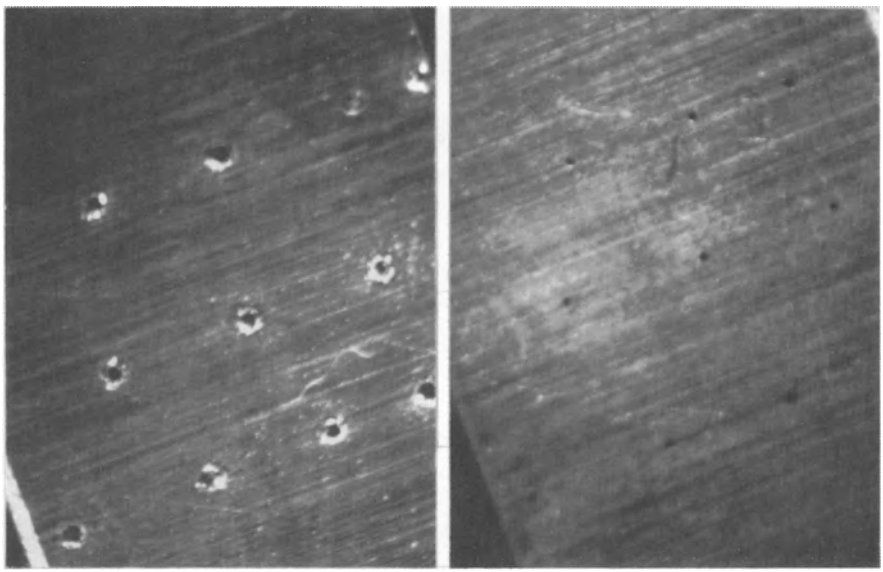


Figure 16-6 Matrix of holes in 0.025-in.-thick stainless steel. The holes were drilled with a pulsed CO_2 laser operating at 75 W average power. The entrances (left) are of about 0.009 in. diameter and have a rim of resolidified material. The exits (right) are of about 0.006 in. diameter and are clean. (From J. F. Ready, *Effects of High-Power Laser Radiation*, Academic Press, New York, 1971.)

Despite these limitations, laser drilling offers a number of advantages that make it useful in many practical hole drilling operations, particularly when the sides of the hole do not have to be exactly straight. The advantages include the following:

- Holes can be drilled with a high throughput rate, leading to low-cost processing.
- The process is easily automated. This also contributes to low cost.
- There is no tool wear. Thus, one major cost associated with mechanical drilling is eliminated.
- One can drill hard, brittle materials, like ceramics and gemstones, which are hard to drill by other means.
- The problem of cracking when drilling ceramics is reduced, so that yields can be high.
- There is no contact of any material with the workpiece, and thus no contamination.
- The heat-affected zone around the hole is very small.
- One can drill extremely small holes in thin materials, smaller than can be produced by other techniques.

Because of these factors, laser drilling of holes has reached production status in industry for many materials, both metallic and nonmetallic.

Many of the applications for laser hole drilling involve nonmetallic materials. Typically, more material may be removed from a nonmetallic material than from a metal by a laser pulse with a given energy, at least partly because of the difference in thermal conductivity. Organic materials are especially easily vaporized. The CO_2 laser is often used with nonmetallic materials because many such materials have high absorption at that laser's wavelength. A pulsed CO_2 laser with an average power around 100 W can effectively drill holes in many nonmetallic materials with a high throughput. Nd:YAG lasers are effective for drilling nonmetallic materials that have high absorption at its wavelength.

Lasers can drill narrow, deep holes through relatively thick brittle materials, like ceramics and silicon. Holes in alumina ceramic with aspect ratios exceeding 25 have been produced. This is a greater aspect ratio than is available with conventional techniques for small holes in ceramics. Ceramic as thick as 0.125 in. can be pierced by repeated laser pulses. Holes may easily be produced close together in patterns or near the edge of a chip of brittle material, operations that are difficult to perform with mechanical drilling. Laser drilling also eliminates costly problems such as drill bit breakage and wear.

One significant application of laser hole drilling that reached production status very early, as early as the 1960s, involved drilling of holes in gemstones. Holes in diamonds used as dies for extrusion of wire are routinely produced by lasers. This has proven a cost-effective application. The hole drilling had previously been very lengthy and costly, because of the hardness of the diamond. Now, the holes can be produced quickly by one or a few pulses from a laser. To make the sides of the holes

smooth, final finishing is still required, but the total time to produce a hole has been drastically shortened. This is an application where laser drilling offered a significant advantage over competing techniques. In another application involving gemstones, laser drilling of holes in ruby watch jewels has become standard practice.

Perhaps the largest application of laser hole drilling is in the drilling of ceramics. This is widely used in the electronics industry. Pulsed CO_2 lasers are widely used to drill holes in alumina substrates for circuits. More specific applications will be given in Chapter 17. Here, we will discuss some of the basic features of ceramic drilling. CO_2 lasers are used for ceramics, because ceramics tend to have high absorption at that laser's wavelength. Laser drilling of hard, high-temperature-fired alumina ceramic is attractive because mechanical drilling has difficulties. Mechanical drilling requires expensive diamond-tipped hardened-steel drill bits. Small holes, less than 0.25 mm in diameter, are difficult to drill mechanically. Breakage of the drill bit often results when the thickness of the ceramic is greater than the diameter of the hole. Also, fracturing of the ceramic limits the yield. Laser drilling easily produces small holes in the brittle material without drill wear and without danger of fracturing the ceramic.

Because of the brittleness of ceramics, conventional hole drilling is often performed on the material in the "green" state before it is fired. But when the ceramic is fired, the dimensions of the part may change and the position of the hole may shift slightly. Laser drilling is performed after firing and thus eliminates this problem.

When a pulsed CO_2 laser drills alumina with thickness around 0.025 in., typical hole diameters may be around 0.005–0.010 in., with tolerance around 0.0005 in. For smaller holes, a Nd:YAG laser will be better because of its shorter wavelength. Frequency-doubled Nd:YAG lasers operating at 0.532 μm have been used for very fine drilling. Very small holes are produced only with some lack of uniformity. For drilling of ceramics, the laser pulse duration is often in the 300–700 μsec range.

Laser drilling is highly compatible with automation. Automated laser systems that drill complicated patterns of holes in material up to 0.6 mm thick have become common. A typical system includes fixturing to hold the ceramic, stepping motors to move the part to predetermined positions at which holes are to be drilled, and a small computer (a numerical controller) to control the laser output and pulse sequencing and to control the positioning of the part. The operator simply inserts a ceramic plate and starts the operation. The entire process then runs under the control of the computer to yield the desired pattern of holes in the ceramic. Figure 16-7 shows an example of such a pattern of laser-drilled holes in an alumina ceramic substrate. The entrances of the holes are clean, and the holes are uniform. The versatility of the laser hole drilling process is apparent.

Laser drilling of ceramics has become very common, more common than laser drilling of metals, because of the limitations described previously (taper, recondensed material, etc.). Also, conventional methods for drilling metals are more satisfactory than are those for drilling ceramics. For ceramics, which are hard, brittle, fracture prone, and which break or wear out drill bits quickly, mechanical drilling is less satisfactory. Laser drilling of such materials provides many advantages and has been one of the biggest successes in the laser material removal field.

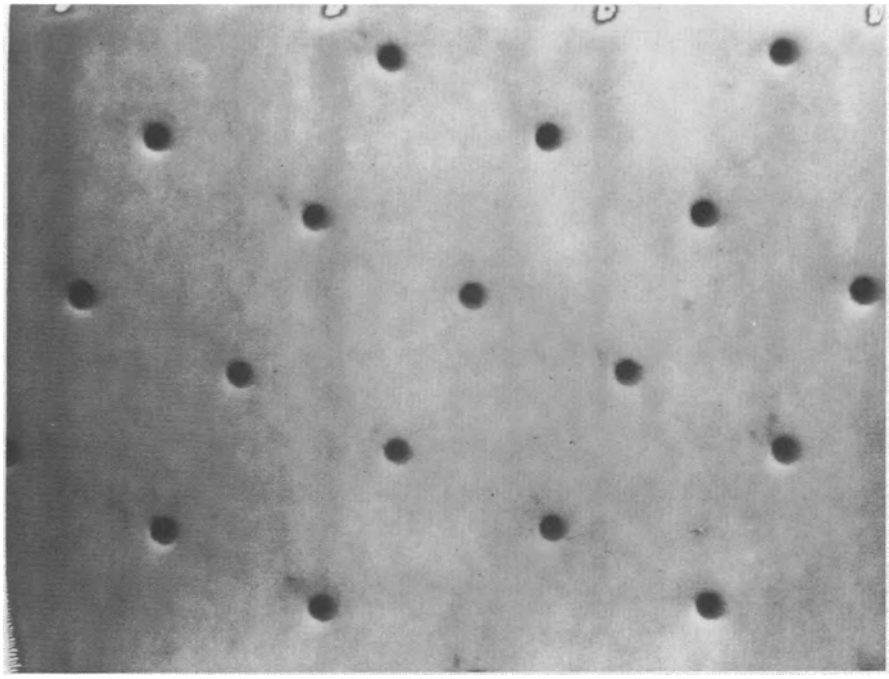


Figure 16-7 A pattern of laser-drilled holes in an alumina ceramic substrate. The diameter of the holes is 0.3 mm.

To summarize, laser hole drilling competes for industrial applications with traditional drilling methods (mechanical drilling) and nontraditional methods, most notably electrical discharge machining (EDM), in which electrically conductive material is removed by repetitive spark discharges. Table 16-3 compares some of the advantages and disadvantages of these three drilling technologies. Laser drilling offers lowest cost when very large numbers of holes are to be drilled. Despite the higher initial equipment cost, the high throughput, easy fixturing, and low operating costs make laser drilling economically favorable. When the limitations associated with laser drilling (hole taper, limited depth and diameter) are important, one of the other techniques will be chosen.

C. Cutting

Lasers may be used to cut or shape a wide variety of materials. By cutting, we mean vaporization of material along a line, so as to separate the workpiece into separate parts. Separation by means of scribing or fracturing will be discussed later. Cutting at practical speeds often involves the use of a gas jet containing oxygen, in order to increase the cut rate. Such oxygen-assisted cutting greatly enhances the capabilities of laser vaporization. Table 16-4 indicates the situations in which different approaches to laser cutting may be employed. Materials with low thermal conductivity

Table 16-3 **Comparison of Drilling Technologies**

	Laser drilling	Mechanical drilling	Electrical discharge machining
Advantages			
	High throughput	Large depth	Large depth
	Easy fixturing	Large diameter	No taper
	No drill wear/breakage	Straight holes	Low equipment cost
	Noncontact process	Low equipment cost	
	Small heat-affected zone		
	Low operating cost		
	Wide range of materials		
Disadvantages			
	Hole taper	Drill wear/breakage	Slow drill rate
	Limited depth/diameter	Low throughput	Long setup time
	Recondensed material	Long setup time	High operating cost
	High equipment cost	Minimum hole diameter	Limited range of materials
		Limited range of materials	

Table 16-4 **Approaches to Laser Cutting**

Material	Approach	Example
Low conductivity	Direct vaporization by laser energy; possibly with nonreactive gas jet	Cloth
High conductivity, nonreactive	Melt with laser energy, flush molten material with gas jet	Stainless steel
High conductivity, reactive	Heat with laser to trigger exothermic reaction with oxidizing gas	Titanium

and low heat of vaporization, like many organic materials, may be burned through directly by the laser without expenditure of too much laser energy. This approach makes the greatest demands on laser power, and it is most applicable to easily vaporized materials. A nonreactive gas jet may be useful to reduce charring. For metals that are not highly chemically reactive (steels), the laser may supply the latent heat of fusion to melt through the material, and then the molten material may be blown away by a gas jet to separate the pieces. In this way, the energy requirements are reduced because one need not supply the latent heat of vaporization. For reactive metals (titanium), the laser may heat the material to the point where an exothermic chemical reaction with an oxidizing gas will begin. The material is then burned through by the chemical reaction. This reduces requirements on laser power still further, because the chemical reaction supplies most of the energy.

Most cutting operations have used CO_2 lasers because of their capability for producing high values of average power. But the advent of multikilowatt Nd:YAG lasers has led to their application in cutting also.

Cutting of nonmetallic materials proceeds easily in many cases. Because of higher reflectivity and higher thermal conductivity, cutting of metals usually requires more laser power than cutting of nonmetals. We first describe cutting of nonmetallic materials and then discuss cutting of metals.

1. CUTTING OF NONMETALLIC MATERIALS

Many nonmetals are cut directly by vaporization of the material, the first entry in Table 16-4. The CO_2 laser is the most commonly employed laser for such cutting. One particularly favorable combination is the cutting of organic materials by continuous CO_2 lasers. Organic materials practically all have high absorption at 10.6 μm, whereas many organic materials are somewhat transparent at the shorter Nd:YAG laser wavelength. Because it requires relatively low energy per unit volume to vaporize most organic materials, the cutting rates can be high.

A continuous CO_2 laser emitting 100 W of power is adequate for many such cutting applications. Materials such as paper, rubber, plastics, cloth, and wood, as well as inorganic materials such as glass and ceramic, can be cut effectively. In one example, slitting of 1/16-in.-thick acrylic sheet has been demonstrated at a rate of 84 ft/min with a CO_2 laser in an operation that leaves the edges clean and slightly fire-polished. In another example, cutting of lightweight paper at a rate greater than 5 m/sec yields edges of quality comparable to a mechanical cut.

Laser cutting of flammable materials is prone to leave ragged edges from scorching. To produce a clean, unscorched edge in laser cutting of such substrates, one may have to blow a jet of inert gas around the cut.

Advantages of CO_2 laser cutting include lack of tool wear, reduced loss of material around the cut, sometimes improved edge quality, ease of automation, simple fixturing, and ability to cut complex shapes easily. The beam may be scanned over the surface under the control of a computer to cut a complex pattern. In this way, the time between the design of a new part and its first fabrication may be substantially reduced. Moreover, laser cutting of materials such as cloth, carpeting, and paper can be economically competitive with mechanical cutting.

The ease of fixturing is especially noteworthy. A sheet of material to be cut is held in place by a vacuum chuck while the laser beam traverses over the surface in a controlled pattern. This simplicity contributes to rapid turnaround time for making new part designs, and it is a strong contributing factor to the economic advantages that laser cutting often offers.

Lasers have been widely used for cutting of plastics. The resulting edge quality may vary greatly, depending on the exact composition of the material and the parameters of the cutting operation. Table 16-5 summarizes how the edges of some common plastics typically respond to the laser cutting operation. A number of plastics can be cut well, but others often yield discolored or charred edges. Thermosetting

Table 16-5 **Cutting of Plastics**[a]

Plastics that are easily cut with good edge quality:

- Acrylic
- Polyethylene
- Polypropylene
- Polystyrene

Plastics that are subject to edge deterioration, ranging from mild discoloration to burning, depending on the exact parameters of the cutting operation:

- Cellulosics
- Fluoropolymers
- Nylon
- Polyesters
- Silicones
- Polyurethane

Plastics for which the edge is always charred to some degree:

- ABS
- Epoxy
- Phenolics
- Polycarbonate
- Polyimide
- Polyvinyl chloride

[a]From W. E. Lawson, Laser Processing of Plastics, Optcon '92, Boston, November 15–20, 1992.

plastics are particularly difficult to cut well. For cutting of plastics, one should use a beam with high irradiance, a well-defined edge, and high beam quality.

Figure 16-8 shows data on cutting of plastics with a 375-W CO_2 laser. The cutting rates are high enough to have practical application.

Table 16-6 presents some other examples of cutting rates measured for nonmetallic materials cut by CO_2 lasers. The numbers in the table are a selection of experimental results, obtained under a variety of conditions. They do not necessarily represent optimum values, but they can help define the capabilities of laser cutting.

One example of cutting that reached practical production status is informative. This is a cloth cutter system that cut material in a single ply [4]. The cloth cutting system contained two 350-W continuous CO_2 lasers, an optical system to move and focus the beam, two x–y positioning systems, a minicomputer for control of the operation, an aluminum honeycomb cutting surface with vacuum-hold down and exhaust, and an automatic cloth spreader. This laser cloth cutting system was completely automated. The cloth was spread on the cutting surface, moved to the desired position under the laser, and cut by the focused laser beam. The laser was stationary, and the beam was moved along the perimeter of the desired pattern by a series of lenses and mirrors. After the cloth was cut, a conveyer moved the cut pieces onto a pickup device. Cutting velocities up to 40 in./sec were possible, but the average speed was generally less. The material was cut in a single thickness. Laser cut-

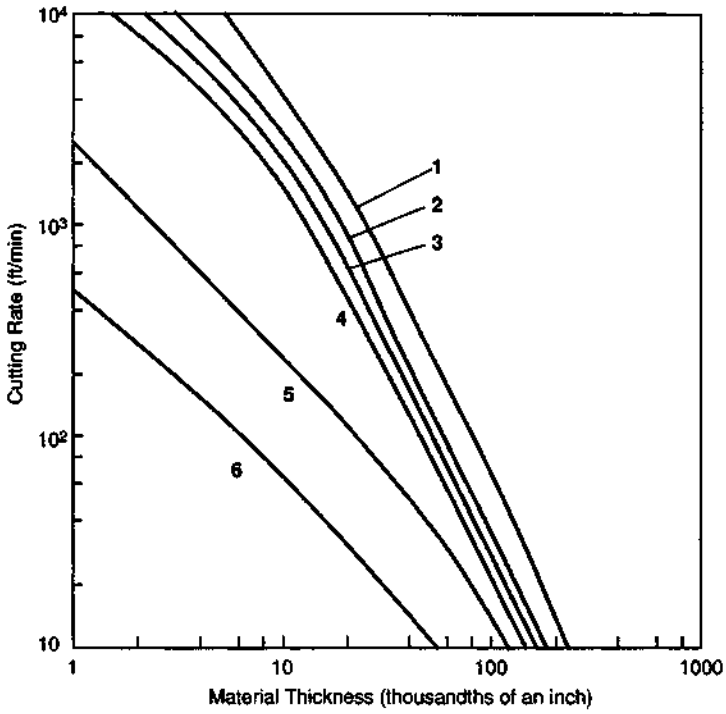


Figure 16-8 Rates for cutting various plastics with a 375-W continuous CO_2 laser, as a function of material thickness: curve 1, cellulose; curve 2, acrylic epoxy; curve 3, ABS plastic; curve 4, vinyl; curve 5, polycarbonate; curve 6, polyethylene. (Data from Photon Sources, Inc.)

ting applied advanced automated technology in an industry that has typically been labor-intensive.

Laser cutting has also been successfully applied in production to making ruled dyes. Slots are cut in wood to accommodate the steel rules used in cutting or creasing cardboard cartons. A typical die may be of 3/4-in.-thick maple wood. Laser cutting of a complicated pattern to hold the steel rules may be completely automated. The cuts are straighter and more uniform than can be produced by a human worker with a jig saw. This technique provided a new automated tool for the production of cartons that are widely used in the food industry.

Another successful application of laser cutting has been separation of alumina substrates into patterns with complex shapes. (Simple separation of alumina into rectangular pieces with straight line cuts may perhaps be performed more simply by scribing, to be described in the next section.) Many integrated circuit applications require alumina to be cut in a complicated shape to serve as a substrate. Conventional methods have been slow and expensive. But sawing the substrates with a CO_2 laser has allowed cutting of odd forms from standard rectangular blanks, with rapid turnaround time and substantial cost savings.

Table 16-6 **Cutting of Nonmetallic Materials with CO_2 Lasers**

Material	Thickness (in.)	Laser power (W)	Cutting rate (in./min)	Gas assist	Reference
Soda lime glass	0.08	350	30	Air	*a*
Quartz	0.125	500	29	Not stated	*b*
Glass	0.125	5,000	180	Yes	*c*
Ceramic tile	0.25	850	19	Yes	*d*
Alumina ceramic	0.024	250	28	Air	*e*
Plywood	0.19	350	209	Air	*a*
Plywood	0.25	850	213	Yes	*d*
Hardboard	0.15	300	36	Air	*a*
Hardboard	0.19	850	180	Yes	*d*
Synthetic rubber	0.1	600	189	None	*d*
Plywood	1	8,000	60	None	*f*
Plywood	0.75	200	12	Air	*g*
Plywood	0.71	500	28	None	*g*
Glass	0.008	350	31	Air	*g*
Glass	0.375	20,000	60	None	*f*
Boron epoxy composite	0.32	15,000	65	None	*f*
Fiberglass epoxy composite	0.5	20,000	180	None	*f*
Acrylic plate	0.04	50	67	Nitrogen	*h*
Acrylic plate	0.22	50	12	Nitrogen	*h*
Plexiglas	0.2	500	110	None	*g*
Plexiglas	0.8	500	14	None	*g*
Graphite epoxy fabric	0.02	400	90	Not stated	*i*
Kevlar epoxy	0.093	300	212	Not stated	*i*

[a] J. E. Harry and F. W. Lunau, *IEEE Trans. Ind. Appl.* **IA-8**, 418 (1972).

[b] B. Feinberg, *Mfr. Eng. Development* (December 1974).

[c] G. K. Chui, *Ceramic Bull.* **54**, 515 (1975)

[d] Data from Ferranti, Ltd.

[e] J. Longfellow, *Solid State Tech.*, p. 45 (August 1973).

[f] E. V. Locke, E. D. Hoag, and R. A. Hella, *IEEE J. Quantum Electron.* **QE-8**, 132 (1972).

[g] D. Schuöcker, Laser Cutting, in *The Industrial Laser Annual Handbook—1986 Edition* (D. Belforte and M. Levitt, eds.), PennWell Books, Tulsa, OK, 1986.

[h] D. Appelt and A. Cunha, in *Laser Technologies in Industry* (O. D. D. Soares, ed.), SPIE Proc. Vol. 952, Part 2, SPIE, Bellingham, WA, 1989.

[i] W. E. Lawson, Laser Cutting of Composites, at the Conference on Composites in Manufacturing 5, Los Amgeles, January 13–16, 1986.

One system [5] used a 250-W continuous CO_2 laser to cut 0.6-mm-thick alumina in complicated contours, at linear velocities of 1.2 cm/sec. A typical pattern could be cut from the parent blank in less than 30 sec, with good detail and excellent edge quality. The cost was lower than for conventional cutting. Moreover, the inventory of shapes needed could be much reduced; one needs only the rectangular blanks.

Because of the ease of fixturing, rapid turnaround time for cutting new designs, and the high throughput rates that are possible, laser cutting of materials such as ceramic, plastic, paper, glass, and cloth has found extensive application in industry.

2. *CUTTING OF METALS*

Laser cutting of metals requires higher average power than the cutting of nonmetals. The higher conductivity and the higher reflectivity of metal surfaces lead to this requirement. Because of the power requirements, most metal cutting applications have been performed with CO_2 lasers.

Oxygen-assisted cutting also is used more often with metals for the same reason. Use of an oxidizing gas can enhance the cutting rate by 40 percent or more compared with the rate when an inert gas is used. Significant thickness of metal may be cut rapidly with a few hundred watts of laser power.

In a typical arrangement, the nozzle for the oxygen (or air) jet is coaxial with the laser beam. The CO_2 laser beam is focused by a suitable infrared-transmitting lens (e.g., germanium or zinc selenide) through the nozzle and onto the workpiece. Typically, a tapered nozzle with 0.05–0.10 in. diameter is used with oxygen supplied at pressures of 15–30 psi.

Figure 16-9 presents data on the cutting rate as a function of laser power for carbon steel with various thicknesses. The cutting rates are high enough to be of practical industrial significance, even with only a few hundred watts of laser power.

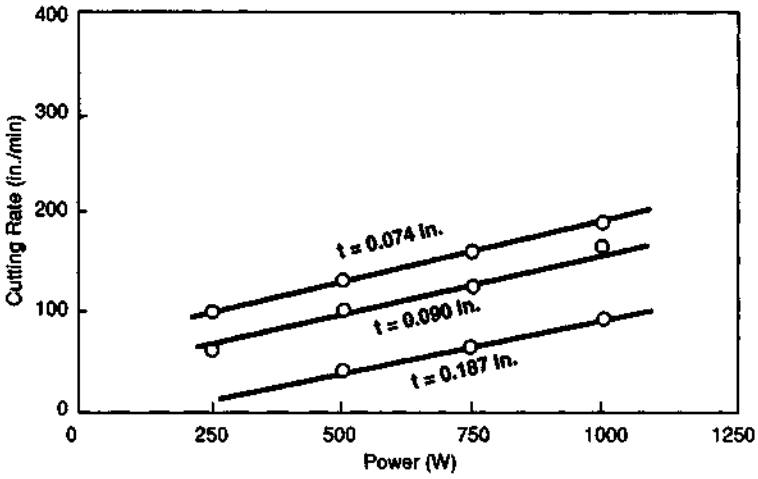


Figure 16-9 Cutting rate for low carbon steel (thickness t) with an oxygen assist. (From S. L. Engel, Society of Manufacturing Engineers Technical Paper MR74-960, 1974.)

With oxygen-assisted cutting, there is a lower limit for the cutting rate below which the material burns excessively, and an upper limit imposed by the available laser power. Figure 16-10 shows data on the lower and upper limits for cutting of low carbon steel with a 1-kW CO_2 laser, as a function of material thickness. The lower limit represents the minimum rate at which burning of the material did not occur. The figure shows that for sufficiently thin material (<0.12 in.), there exists a reasonable process window for a cutting operation.

Figure 16-11 shows data on the capability of lasers for cutting carbon steel with greater thickness. Data are presented for two levels of laser power, 1000 and 1500 W. The results in these two figures indicate cutting rates high enough to be of practical economic importance.

To present data on a wider range of materials and conditions, Table 16-7 illustrates cutting rates achievable with oxygen-assisted cutting by CO_2 lasers operating at levels of some hundreds of watts, for a variety of metals. The cutting rates would be substantially lower without the oxygen assist. The data in this table, assembled from several sources, are intended to illustrate some representative values for experimental cutting rates and are not necessarily optimized values. Moreover, quoted cutting rates may differ when values from different sources are compared, even for nominally similar conditions. Uncontrolled variables and possible differences in procedures for power measurement may account for the differences. The numbers

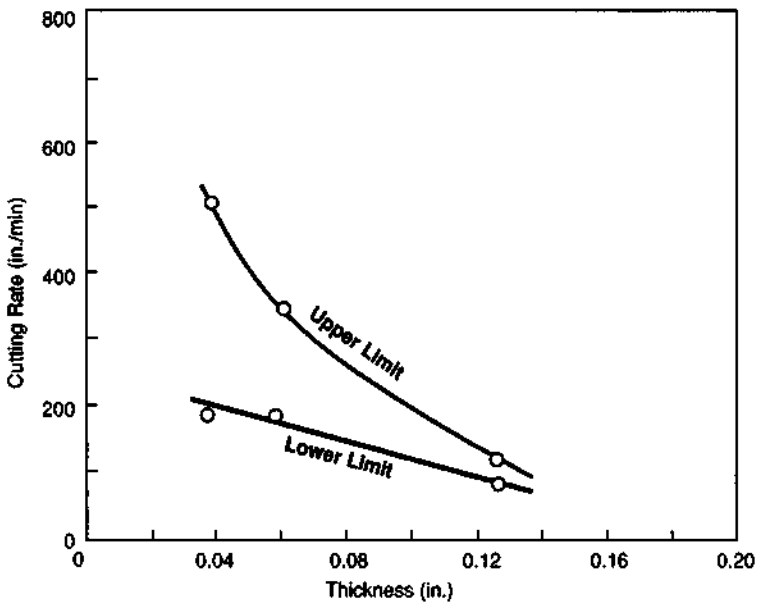


Figure 16-10 Maximum and minimum cutting rates for low carbon steel at 1 kW laser power. (From S. L. Engel, Society of Manufacturing Engineers Technical Paper MR74-960, 1974.)

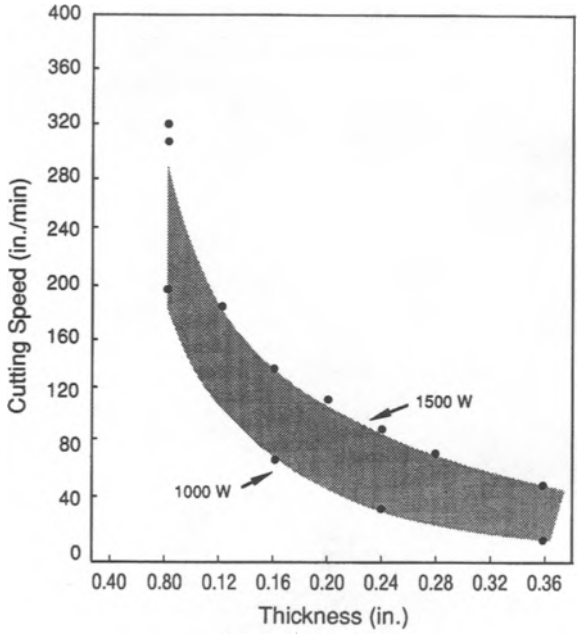


Figure 16-11 Cutting rate as a function of thickness for oxygen-assisted cutting of carbon steel with a carbon dioxide laser, at the indicated levels of laser power. (From *The Industrial Laser Annual Handbook—1990 Edition* (D. Belforte and M. Levitt, eds.), PennWell Books, Tulsa, OK, 1990.)

in Table 16-7 define an approximate range for the capability for oxygen-assisted laser cutting. The cutting rates are high enough to be of commercial interest.

Oxygen-assisted cutting is most suitable for reactive metals, like titanium. The operation yields cuts with narrow heat-affected zones and with small cut width, compared those produced in conventional cutting. Most work reported to date has used high-power continuous or repetitively pulsed CO_2 lasers, but some Nd:YAG laser cutting is now being reported.

As Nd:YAG lasers have reached higher levels of average power in recent years, they also have become candidates for cutting of metals. Figure 16-12 shows results for cutting of several metals with a pulsed Nd:YAG laser. At the same level of laser power, a Nd:YAG laser will generally have greater cutting capability than a CO_2 laser, because the beam can be focused to a finer spot and because the surface reflectivity is lower. Still, because Nd:YAG lasers cannot reach the same high levels of power as CO_2 devices, they have not been used for as many cutting operations as CO_2 lasers.

The cutting capabilities of multikilowatt lasers are even greater than what has just been described. An oxygen assist may no longer be needed. The very high levels of power available may supply the necessary energy to produce acceptable cutting at an economic rate. In fact, a nonreactive shield gas (like helium) may need to be used

Table 16-7 Oxygen-Assisted Cutting of Metals with CO_2 Lasers

Metal	Thickness (in.)	Laser power (W)	Cutting rate (in./min)	Reference
Titanium (pure)	0.02	135	600	*a*
Titanium (pure)	0.67	240	240	*b*
Titanium alloy (6Al4V)	0.05	210	300	*a*
Titanium alloy (6Al4V)	0.088	210	150	*a*
Titanium alloy (6Al4V)	0.25	250	110	*a*
Titanium alloy (6Al4V)	0.39	260	100	*a*
Titanium alloy	0.2	850	130	*c*
Carbon steel	0.125	190	22	*a*
Stainless steel (302)	0.012	200	90	*d*
Stainless steel (410)	0.063	250	50	*e*
Stainless steel (410)	0.11	250	10	*e*
Stainless steel	0.012	350	170	*b*
Galvanized steel	0.039	400	177	*f*
Nickel alloy Nimonic 75	0.03	200	16	*g*
Rene 41	0.02	250	80	*e*
Rene 41	0.05	250	20	*e*
Zircalloy	0.018	230	600	*a*
Molybdenum	0.002	500	16	*f*
Niobium	0.126	500	8	*f*

[a]Booklet entitled *CO_2 Applications*, Coherent Radiation, Inc., Palo Alto, CA, 1969.

[b]J. E. Harry and F. W. Lunau, *IEEE Trans. Ind. Appl.* **IA-8**, 418 (1972).

[c]Data from Ferranti, Ltd.

[d]W. W. Duley and J. N. Gonsalves, *Opt. and Laser Technol.*, p. 78 (April 1974).

[e]J. R. Williamson, in *Industrial Applications of High Power Laser Technology* (J. F. Ready, ed.), SPIE Proc., Vol. 86, SPIE, Palos Verdes Estates, CA, 1976.

[f]*Guide to Laser Materials Processing*, Laser Institute of America, Orlando, FL, 1993.

[g]D. Schuöcker, Laser Cutting, in *The Inductrial Laser Annual Handbook—1986 Edition* (D. Belforte and M. Levitt, eds.), PennWell Books, Tulsa, OK, 1986.

to avoid burning. The thickness that can be cut and the cutting speeds are impressive. A tabulation of published values for various cases of multikilowatt CO_2 laser cutting appears in Table 16-8. The numbers do not necessarily represent optimum values, but are indicative of what can be achieved.

An estimate for the maximum thickness of mild steel for which oxygen-assisted cutting is possible is shown in Figure 16-13 as a function of laser power. At a level around 10 kW, one can cut material with thickness in excess of 2 in. The line represents the maximum possible depth that can be cut with present technology.

Laser cutting of metals can be an extremely versatile process. Complex patterns may easily be cut. Figure 16-14 illustrates this. It shows an example of cutting a pattern of square holes in 0.02-in.-thick stainless steel with a CO_2 laser. The laser

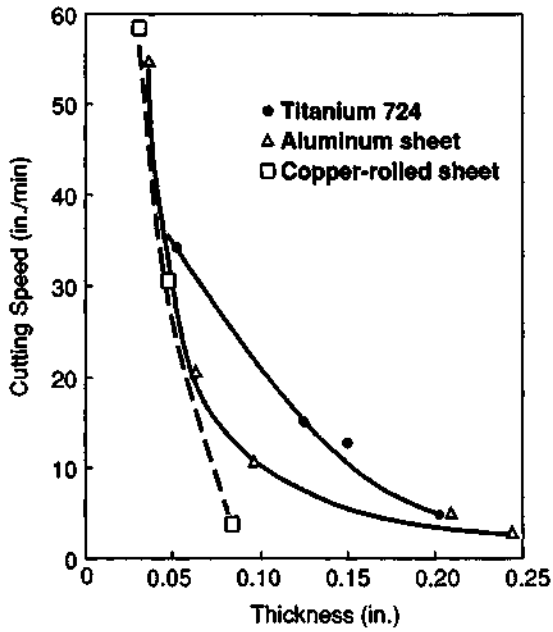


Figure 16-12 Cutting rate as a function of thickness for oxygen-assisted cutting of various metals by a 400-W pulsed Nd:YAG laser. (From *The Industrial Laser Annual Handbook—1990 Edition* (D. Belforte and M. Levitt, eds.), PennWell Books, Tulsa, OK, 1990.)

cutting of this pattern is easy, especially when the simple fixturing requirements are considered. Cutting of such patterns by other processes is more difficult.

To obtain good-quality reproducible results with laser cutting, one must control the polarization of the laser beam. One obtains the best results when the laser beam is linearly polarized in the same direction as the direction of travel for the cut. In this case, the cutting depth is maximized and the kerf is minimized. The edges of the cut are straight and perpendicular to the surface of the metal. But if the direction of polarization differs from the direction of travel, the edge quality deteriorates, the cut widens, and the penetration decreases. The edges of the cut become curved and are no longer perpendicular to the surface. The overall quality of the cut decreases substantially. The results are worst when the polarization is at right angles to the cut direction. Thus, one obtains the best results when the polarization is controlled so as to be parallel to the cut direction.

This may be difficult to accomplish when one is cutting complex shapes, because it is impractical to change the polarization continuously as the cut direction changes. In practice, one uses a laser with a circularly polarized beam. This is accomplished by inserting a quarter-wave plate into the linearly polarized beam. The circularly polarized beam produces a cut with high quality, as good as the results with a beam linearly polarized in the direction of travel. Thus, use of a circularly polarized beam

Table 16-8 Metal Cutting with Multikilowatt CO_2 Lasers

Metal	Thickness (in.)	Power (kW)	Cutting rate (in./min)	Reference
Aluminum (6061)	0.5	10	40	*a*
Aluminum	0.04	3.0	250	*b*
Aluminum	0.125	4.0	100	*b*
Aluminum	0.25	3.8	40	*b*
Aluminum	0.50	5.7	30	*b*
Aluminum alloy	0.5	12	100	*c*
Aluminum	0.5	15	92	*d*
Inconel	0.5	11	50	*a*
Stainless steel (300)	0.125	3.0	100	*b*
Steel (low carbon)	0.125	3.0	100	*b*
Steel (low carbon)	0.66	3.0	45	*b*
Steel (1018)	0.5	12	60	*c*
Steel (1018)	0.75	12	20	*c*
Steel (304)	0.5	12	60	*c*
Steel (304)	1.0	15	20	*c*
Titanium	0.25	3.0	140	*b*
Titanium	1.25	3.0	50	*b*
Titanium	2.0	3.0	20	*b*
Titanium	1	10	204	*d*
Titanium alloy	0.7	13.5	35	*c*
Titanium alloy	2.0	13.5	6.5	*c*
Rene 95	1.25	18	15	*c*
Rene 95	2.2	18	2.5	*c*

[a]E. V. Locke and R. A. Hella, *IEEE J. Quantum Electron.* **QE-10**, 179 (1974).

[b]D. W. Wick, Society of Manufacturing Engineers Technical Paper MR75-491, 1975.

[c]S. S. Charschan, ed., *Guide to Laser Materials Processing*, Laser Institute of America, Orlando, FL, 1993.

[d]D. Schuöcker, Laser Cutting, in *The Industrial Laser Annual Handbook—1986 Edition* (D. Belforte and M. Levitt, eds.), PennWell Books, Tulsa, OK, 1986.

for cutting is the practical solution for obtaining good cutting results without introducing unnecessary complexity into the system.

Another recent innovation that improves the quality of the cut edge involves the use of a nitrogen jet at high pressure, greater than 200 psi. Such high-pressure cutting gives a bright clean edge to metals cut by a laser. The amount of hardened dross at the side of the cut is reduced, so that less post-cut cleaning is required. High-pressure cutting has proved to be useful for a variety of metals, including stainless steel, aluminum, and brass.

To summarize the discussion of laser cutting of metals, we conclude with comments on the suitability of various types of metals for laser cutting, on the costs of laser cutting, and on the comparison of laser cutting with other technologies.

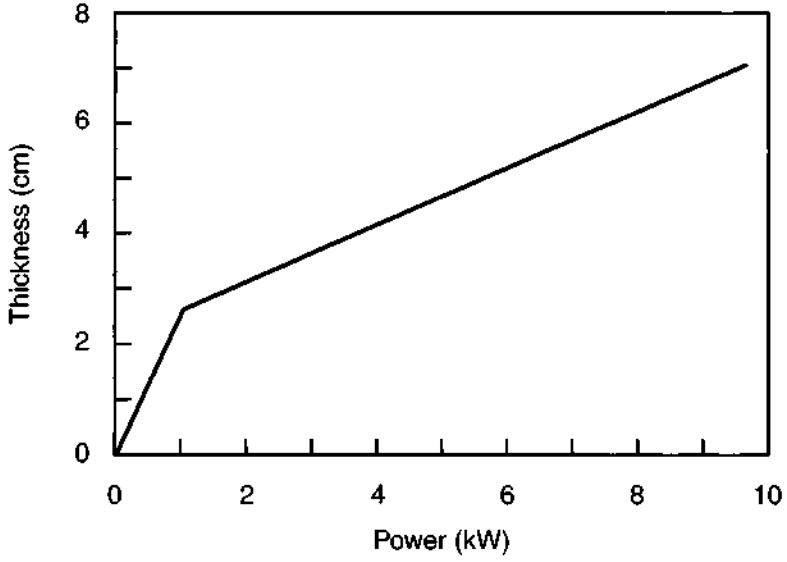


Figure 16-13 Maximum cut depth for oxygen-assisted laser cutting of mild steel as a function of laser power. The solid line is an estimate of the maximum cut depth that may be achieved. (From C. Williamson, G. E. Fox, and V. G. Gregson, at the International Conference on Applications of Lasers and Electro-Optics, Santa Clara, CA, October 30–November 4, 1988.)

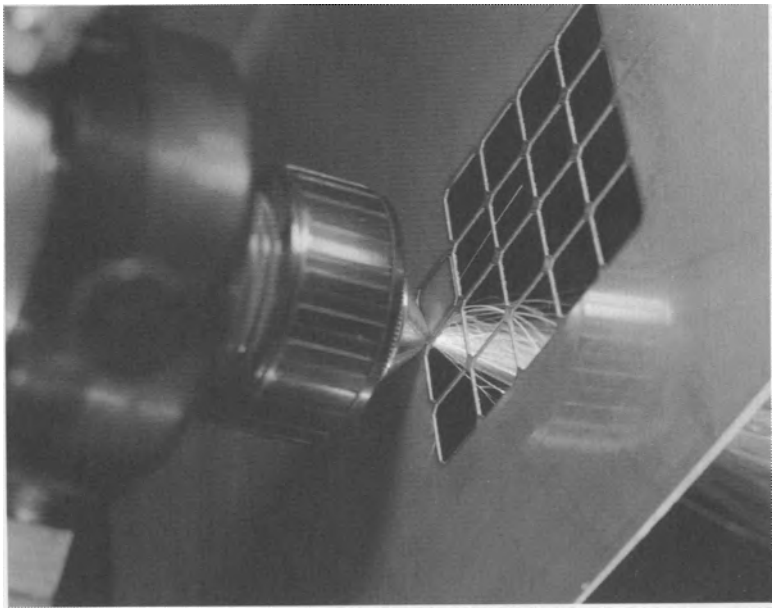


Figure 16-14 Laser cutting of stainless steel. Square openings are being cut in 0.02-in.-thick steel with a high-power CO_2 laser. (Photograph courtesy of S. L. Engel and GTE Sylvania.)

Table 16-9 compares the quality of laser cutting for several metals for both Nd:YAG and CO_2 lasers. The quality of the cut is good to excellent for many metals. For ferrous metals, CO_2 lasers give excellent results, but for metals with high infrared reflectivity (like aluminum), Nd:YAG lasers seem to give better results. The implication of this table is that high-quality laser cutting may be obtained with many metals.

As an example, laser cutting has reached production status in the aerospace industry for cutting of metals like titanium. Carbon dioxide laser systems operating at levels of a few hundred watts have been suitable for cutting of hundreds of thousands of feet of titanium per year [6]. Sophisticated multiaxis systems are employed for cutting contoured surfaces for airframes. Despite extensive study, cutting of aluminum has apparently not yet reached production status in the aerospace industry.

The ease of fixturing, speedy setup, and ease of automation make laser cutting economically attractive, even in competition with technologies with lower capital costs. Oxygen-assisted laser cutting offers considerable savings for certain types of metal, because of reduced material loss and reduced machining required to clean the cut. One economic analysis [6] indicated substantially lower costs for laser cutting of titanium compared with costs for cutting with a band saw, even though the initial cost of the automated laser system was substantially higher.

Laser cutting competes with many other technologies in industrial applications. The other technologies include both traditional and nontraditional machining methods. The traditional methods, like the use of a punch press, are mainly well-established mechanical methods. The term "nontraditional machining" is applied to material removal processes that have recently emerged, like plasma arc cutting and wire electric discharge machining, as well as laser cutting. Each cutting technology has its own set of advantages and disadvantages, which are summarized in Table 16-10 for a selection of both traditional and nontraditional cutting technologies.

Table 16-9 **Quality of Laser Cutting of Various Metals**[a]

Metal	CO_2 laser	Nd:YAG laser
Aluminum (1000 series)	Fair–good	Good
Aluminum (2000 series)	Good	Good–excellent
Brass	Fair–good	Fair–good
Carbon steel (low)	Excellent	—
Carbon steel (medium)	Excellent	Fair–good
Carbon steel (high)	Good	Good–excellent
Copper	Fair–good	Good–excellent
Inconel	Good	Good–excellent
Stainless steel (304)	Good–excellent	Good
Stainless steel (316)	Good–excellent	Good–excellent
Titanium	Good	Good

[a]Based on information from *The Industrial Laser Annual Handbook—1990 Edition* (D. Belforte and M. Levitt, eds.), PennWell Books, Tulsa, OK, 1990.

Table 16-10 **Comparison of Cutting Technologies**

Technology	Advantages	Disadvantages
Laser cutting	Ease of fixturing Fast setup and rapid design change Small kerf Capability to cut complex shapes easily	High capital cost Relatively slow removal rate High power requirements
Punch press	High rate and low cost for volume applications	High die and tooling costs Long design and setup times
Numerically controlled milling	Good tolerances Good edge quality High material removal rate	Long setup time Tool wear
Abrasive fluid jet	Good tolerance Capability for complex shapes Good surface finish Cuts almost all materials	High equipment cost Relatively slow cutting
Wire electric discharge machining	Good edge quality Can cut thick metals	Fixturing Electrode wear
Plasma arc	Cuts thick material well High cutting rate Capability for complex shapes	High operating cost Poor tolerance Large kerf

Because each technology offers advantages that may be useful in specific applications, all the technologies will remain in use, for the applications that favor them. Laser cutting has established itself as a viable competitor for many applications, on both a technical and an economic basis. It has become the most widely used industrial application of laser processing.

D. Scribing

Laser scribing is a method for shaping and separating a workpiece into segments of desired shape without having to vaporize all the way through the material. It may be regarded as an alternative to cutting. It is often used to dice wafers of material into smaller chips. Scribing is usually applied to shaping of brittle materials such as ceramics, silicon, and glass. It is usually performed by drilling a series of closely spaced blind holes into the surface of the material. Typically, the hole depth might be one-quarter of the thickness of the workpiece. The material will then snap easily along the line described by the scribing. If a wafer is scribed with a series of lines in

two perpendicular directions, it may easily be separated into a large number of small chips. The quality of the edge defined by the scribing and snapping is of sufficient quality for many applications. The chips are then ready for further processing or packaging.

Scribing has also been performed by etching a continuous groove into the surface. The drilling of a line of blind holes, though, has more often been employed.

The CO_2 laser is usually used for scribing ceramics, because of their high absorption at 10.6 μm. Both continuous and repetitively pulsed lasers have been used. Table 16-11 presents typical parameters for CO_2 laser scribing of alumina ceramic. If silicon is to be scribed, its high absorption near 1 μm makes the Nd:YAG laser a better choice. Usually, a Q-switched Nd:YAG laser has been used.

Direct cutting of the wafers is an alternative method for dicing wafers. Compared with direct cutting, laser scribing requires less energy per unit length, because one does not cut all the way through the material. Thus, scribing for straight line separations will be faster than direct cutting. For a typical case of separating alumina of thickness 0.03 in. with a CO_2 laser having a few hundred watts of output power, cutting may proceed at a rate of a few tenths of an inch per second, whereas scribing may proceed at velocities up to 10 in./sec. Also, in comparison with direct cutting, scribing produces edges with less damage due to heating and with less thermal stress. Because of these considerations, scribing is preferred in situations where only straight separations are needed, such as dicing wafers into rectangular pieces. Direct cutting is applicable when sharp corners or complex shapes are required.

As specific examples, silicon wafers have been scribed at 0.6 in./min by a Nd:YAG laser; and 0.025-in.-thick alumina substrates have been scribed at 60 in./min by a 100-W continuous CO_2 laser.

The use of laser scribing has led to significant increases in production rates for the dicing of silicon wafers. The quality of the edges compares well with the results obtained with a diamond saw. The laser beam scribes a straight uniform line of holes with accurately controlled depth. Hundreds of individual chips are obtained by snapping the wafer along the scribe lines.

Laser scribing of materials like silicon and ceramics offers the advantage of clean separations without contamination and without tool wear. Laser scribing has become well established in industrial applications. According to one estimate [7], billions of holes are drilled daily in ceramic scribing operations.

Table 16-11 **Typical CO_2 Laser Parameters for Scribing of Alumina**

Peak power	300 W
Repetition rate	1 kHz
Pulse duration	100 μsec
Hole diameter	0.005 in.
Hole depth	0.007 in.
Scribe speed	6 in./sec

Laser scribing is economically competitive, reducing costs associated with tool wear and breakage, and increasing production yield. One estimate [8] of relative operating costs (neglecting capital expenditures) indicated that the costs for laser scribing were less than 20 percent of the costs associated with diamond scribing.

Another laser-based technique for separation of brittle materials is controlled fracturing. One may fracture brittle materials by sudden stress caused by rapid heating. When laser energy is absorbed at the surface of a brittle material, it heats the material rapidly, causing mechanical stress that can lead to localized fracture. If the beam is moved along the surface of the material, the fracture will follow the beam path. If the fracture can be controlled, it can separate the material without surface damage and without removal of any material.

The technique of controlled fracturing could be used for separating components like microcircuit chips without damage or contamination. In one demonstration, a CO_2 laser with power less than 50 W separated 0.027-in.-thick alumina at a rate over 60 in./min [9]. Separations could be made along controlled paths, not necessarily only straight lines.

In comparison with scribing, fracturing has a problem of possible wandering of the fracture path, particularly if one tries to turn a sharp corner. It has been difficult to control the path of the separation reproducibly. Because of this fact, scribing has more general usefulness. Despite some early enthusiasm for controlled fracturing as a method for separating brittle materials, it does not seem to have been widely adopted as a practical method for separation.

E. Marking

Lasers can fulfill many requirements for marking of manufactured parts. Product marking can be carried out for purposes of identification, for product information, for imprinting distinctive logos, for identification of gemstones, and for theft prevention. Conventional marking techniques include printing, stamping, mechanical engraving, manual scribing, etching, and sandblasting. In some cases, needs for durable permanent marks stretch the capabilities of these techniques.

Laser marking offers a flexible, sophisticated alternative to conventional marking of products. Laser marking offers many potential advantages. Virtually any material can be marked with permanent, high-quality marks. Laser marking is a noncontact process, and thus it minimizes mechanical distortion and introduces no contamination. It is especially applicable to marking of small, delicate, high-value assemblies, like semiconductor wafers. Thus, even though laser equipment is often more expensive than competing technologies, laser marking applications in industry have become common and are increasing in number.

Different approaches to laser marking are employed. In the first method, one scans the laser beam over the surface of the part and modulates it to vaporize a small amount of material at selected positions, thus forming a mark. The mark may be alphanumeric characters or any other desired pattern. Often, the beam is focused to a small area at the surface, and the mark is fabricated as a dot matrix pattern

consisting of many small vaporized pits, formed by many pulses of the laser. The lasers used in this technique are pulsed and have high peak power. The devices most often used are Q-switched Nd:YAG lasers and excimer lasers. This approach is used for hard materials with relatively high latent heat of vaporization, like metals or semiconductors.

When marking is performed in a dot matrix pattern, a series of tiny holes defines the desired characters or figures. The high-power beam is focused on the surface, scanned in the desired pattern, and pulsed when a mark is needed.

The scanning technique has been employed for imprinting identification marks on products such as silicon wafers and expensive typewriters. The Nd:YAG laser has often been employed for this application, because the materials to be marked absorb well at the 1.06 μm wavelength, and because it offers short pulses of high peak power, capable of vaporizing a mark onto the surface.

Laser printing of identifying numbers on silicon wafers is attractive because of the brittleness of the material, because of the lack of contamination, and because later processing operations could destroy printed marks. Laser marking of expensive typewriters offers a distinctive method of identification that cannot be duplicated well by other methods.

In a variation of the scanning approach, commonly called laser engraving, the laser is used to etch grooves into a workpiece. The laser vaporizes the workpiece material and leaves a groove, typically about 0.006 in. wide and with straight vertical walls. The beam may be scanned and modulated so as to engrave characters or markings of any desired configuration.

Laser engraving has often been used to mark wood products for applications like fabrication of artistic scenes or commemorative plaques. This application often employs a CO_2 laser operating in the range of 40–80 W of power. The pattern may be stored in a computer memory, which controls the scanning and the modulation of the beam. Alternatively, the laser may be controlled by the output from a photodetector that scans an original drawing synchronously with the laser scan. This method allows rapid transformation of a drawing into an engraved pattern.

A second broad approach to laser marking, sometimes called image micromachining, forms the pattern all at once in a single pulse of the laser, by delivering a high-power beam already focused in the desired pattern. In this approach, the beam is projected through a mask and focused onto the surface. This approach usually has employed CO_2 lasers. It is most applicable to marking of softer materials with relatively low latent heat of vaporization, like organic materials.

Image micromachining can produce a complicated pattern all in one laser pulse. A schematic diagram is shown in Figure 16-15. The optical arrangement is similar to that of a slide projector. A mask with the desired pattern is inserted in the beam. The lens images the mask pattern on the surface. If the laser power is high enough, surface material is vaporized, and a permanent image of the mark is imprinted.

The emphasis in image micromachining has been to imprint fairly large area patterns in a single pulse. Hence, the peak power must be high. The materials most easily marked in this way are organic materials, like paper and wood. The laser most

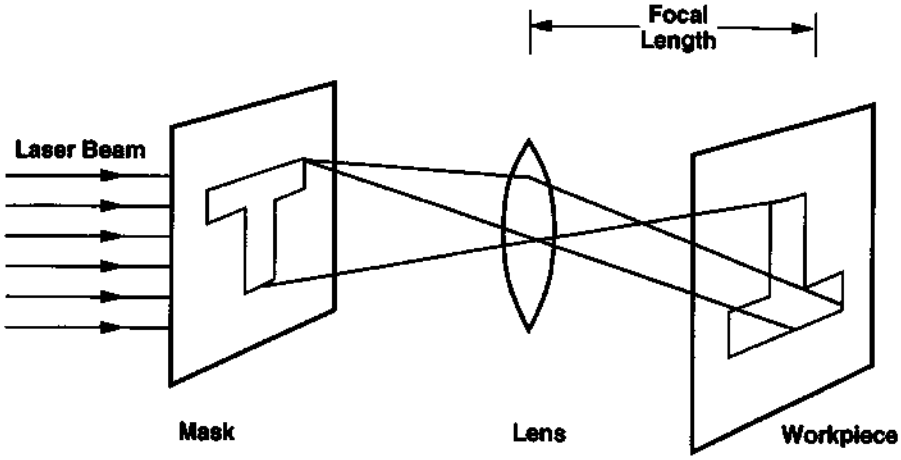


Figure 16-15 Schematic diagram of image micromachining.

often employed has been the CO_2 TEA laser, because the CO_2 laser wavelength is absorbed well by organic materials. A TEA laser of reasonable size can mark an area around 1 cm^2 on typical packaging materials in one pulse. This allows applications like coding for dating or batch identification, or imprinting of specialized logos.

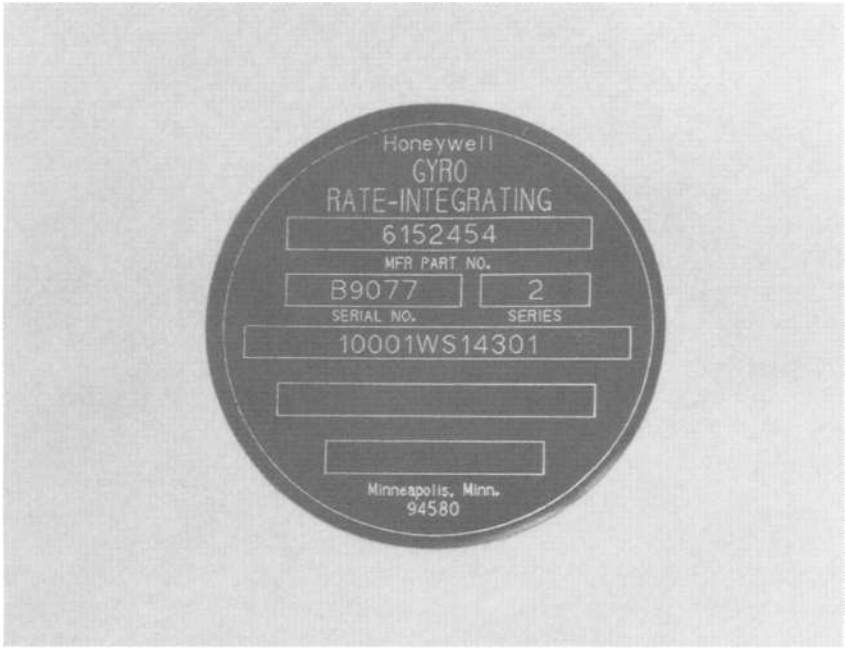
Although CO_2 lasers have been used often for this application, developing excimer laser technology has made those lasers good competitors. The absorption of most materials at the ultraviolet excimer laser wavelengths can be very high, and the short wavelength allows better pattern definition than use of the longer CO_2 laser wavelength. Figure 16-16 shows two examples of laser produced marks, one formed by a dot matrix approach and one by image micromachining.

Numerous laser marking systems are in industrial use, and there appears to be a growing demand for laser marking.

F. Balancing

Dynamic balancing of a workpiece can be performed by a laser while the workpiece remains in motion. Most conventional balancing operations require determination of the heavy area while the workpiece is in motion. Then, one stops the workpiece and mechanically removes material from the heavy region; one then restarts the motion and redetermines the heavy area. This process involves trial and error and often is slow, labor-intensive, and expensive.

Laser pulse durations are short enough that material can be removed from the heavy spot while the workpiece is rotating in the test fixture. The laser pulse can be synchronized with the detection of the heavy portion of the rotating workpiece, so that the laser pulse strikes the heavy region. The unbalance is detected by conventional methods, and each laser pulse is directed at the heavy region until the



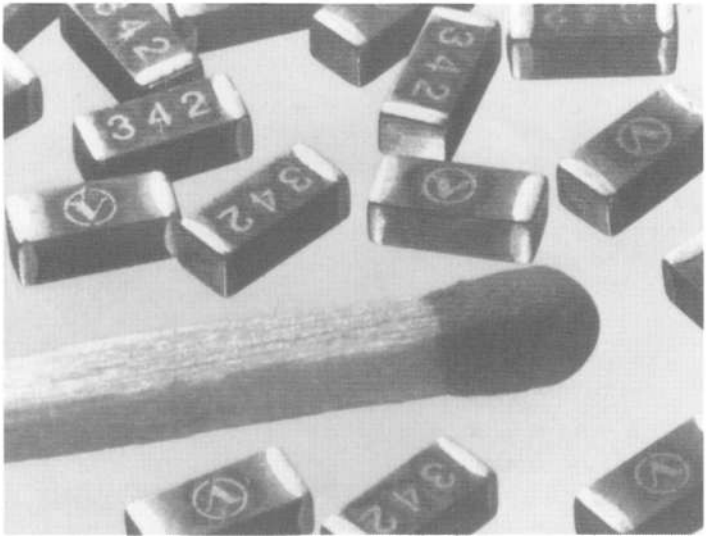


Figure 16-16 Examples of laser marking. Top: Identification marks made in anodized aluminum on a rate-integrating gyroscope in a dot matrix pattern using a CO_2 laser. (Photograph courtesy of Honeywell, Inc.) Bottom: Marking of an electronic package by image micromachining using an excimer laser. (Photograph courtesy of Lambda Physik, Inc.)

workpiece is brought into balance. Because the workpiece does not have to be stopped before removing material, laser balancing can produce substantial savings in time and expense. Balancing to less than 10 μoz-in. of residual imbalance is possible.

A relatively small pulsed laser may remove about 1 mg of material per pulse, an amount adequate for many balancing operations. The pulse duration must be short. Because of the rotation, the energy will be deposited over some angular sector of the workpiece, and the amount of material will be reduced compared with the amount removed from a stationary target. If the laser has a pulse duration around 1 msec, and the rotation rate is high, the energy will be deposited around a large angular sector. Thus, pulse durations of a few tens of microseconds are desirable to keep the energy better localized in the heavy region. With such a pulse duration, rotors can be balanced at rotation rates up to 24,000 rpm. Even at high rotation rates, a laser emitting around 10 J per pulse is adequate for dynamic balancing, provided that the pulse repetition rate is not too low.

Laser balancing has been used to balance a variety of industrial products, including gyroscope rotors, turboprop rotors, and electric motor armatures, and to adjust crystal oscillators for wristwatches.

G. Paint Stripping

An application that has been studied for many years and that has recently been accepted as a practical process is the removal of paint. This process has been developed primarily for use on military aircraft.

The laser provides irradiance high enough to vaporize the paint, but not to affect the underlying surface. Laser stripping produces no damage, in contrast to the former process of chemical stripping, which did create some damage.

In one example, a 5.4-kW CO_2 laser can remove 0.001-in.-thick paint from 1200 ft^2 of surface per hour. It can strip the paint from the radome of a fighter aircraft in 1.5 hours, leaving one cup of solid residue. By comparison, chemical stripping required 8–10 hours and left more than 100 gallons of hazardous waste.

Although not yet a large application, this type of processing illustrates the versatility of laser material removal.

H. Laser Deposition of Thin Films

Laser vaporization has been utilized for deposition of thin films for varying applications. A variety of deposition techniques have been studied, including physical vapor deposition, chemical vapor deposition, and thermally induced deposition. Only physical vapor deposition is properly the subject of this chapter. The other techniques will be discussed in Chapter 17, in conjunction with the fabrication of microelectronic structures.

In a typical deposition experiment, the laser beam, usually from a pulsed laser with high peak power, is directed at a target of the material to be deposited. The

process is usually carried out in vacuum, or at least a partial vacuum. The laser evaporates material from the target surface. The substrate to be coated is located a short distance from the target. Some of the evaporated material condenses on the substrate. This leads to growth of a thin film of the desired material. In some cases, the composition of the film may be varied by chemical interactions with ambient gas in the system.

There have been many demonstrations of the growth of a wide variety of films, including metals, insulators, and semiconductors. But the transition to use of laser deposition for production has been slow.

There has been substantial interest in the use of lasers for deposition of high-temperature superconducting films, such as $YBa_2Cu_3O_{7-x}$. In one example [10], superconducting thin films of $Tl_2CaBa_2Cu_2O_x$ were routinely prepared by laser ablation in an atmosphere of 20 mTorr of oxygen. The laser used was a krypton fluoride device operating at 248 nm. After deposition, the films were annealed. After annealing, x-ray analysis showed only the superconducting phase of the material epitaxially aligned to the $LaAlO_3$ substrate. The critical temperature for the superconducting transition was quite uniform. This one example shows the potential utility of laser ablation for deposition of superconducting films.

I. Specific Examples of Material Removal

One specific example of laser cutting in practical industrial use involves cutting of flat stock patterns for use in fabrication of custom food preparation work areas. The two-dimensional blanks are bent into three-dimensional shapes and then welded into a finished assembly. The finished product is specialized food service lines for restaurants, schools, hospitals, and so on.

The starting blanks, of stainless steel, are cut with a 1.8-kW CO_2 laser beam. The work system allows processing of 50 sheets of 4 by 12 foot stainless steel in one shift. The laser cutting produces cuts that have no burrs nor dross. Laser cutting provides superior accuracy and better fitting than the previous process, plasma cutting. Because of the superior fitting, the subsequent welding process is faster and easier. Because of the high cut quality, additional post-cutting operations are eliminated. The result is lowered cost, simplified design, and higher throughput for the assembly of the food service lines. In addition, the improved fit makes it easier to meet hygiene standards for food preparation equipment.

A second specific example involves fabrication of sheet metal components for aircraft engines. One application is in the trimming of manifold tubes. The tubes are formed by a combination of hydroforming and laser trimming. The starting material is flat inconel sheet, which is formed into an approximate shape on a hydroform press. The excess metal that remains along the edges of the desired final form is trimmed with a 2-kW CO_2 laser. The trimming yields half of a tube bent in the middle like an elbow joint. Two of these half-sections are then welded to form the tube.

The hydroformed part is mounted on a rotary table. During trimming, the part is held stationary and the laser head moves. The laser is a five-axis model that can trim

all the way around a three-dimensional part, under the control of a special purpose CAD/CAM software program.

Previously, the trimming had been performed by manual scribing and sanding. Each half-section required about 20 minutes to complete. The laser trimming requires between 1 and 2 minutes, including loading and unloading. In the laser processing, the part is mounted with simple clamping, which eliminates the need for specialized trim dies. Thus, laser trimming reduced costs substantially for this part, as well as cutting the lead time significantly. It also provides higher part quality than the previous scribing and sanding.

References

[1] M. K. Chun and K. Rose, *J. Appl. Phys.* **41**, 614 (1970).
[2] F. P. Gagliano and U. C. Paek, *Appl. Opt.* **13**, 274 (1974).
[3] H. W. Bergmann *et al.*, paper CWJ3 at the Conference on Lasers and Electro-Optics, Anaheim, CA, May 10–15, 1992.
[4] D. W. Wick, Applications for Industrial Lasercutter Systems, Society of Manufacturing Engineers Technical Paper MR75-491, 1975.
[5] J. Longfellow, *Solid State Technol.*, p. 45 (August 1973).
[6] J. R. Williamson, in *Industrial Applications of High Power Laser Technology* (J. F. Ready, ed.), Vol. 86, Proc. Soc. Photo-Opt. Instrum. Engineers, Palos Verdes Estates, CA, 1976.
[7] R. Barber, Society of Manufacturing Engineers Technical Paper MR74-951, 1974.
[8] *MicroWaves*, p. 71 (August 1970).
[9] R. M. Lumley, *Ceramic Bull.* **48**, 850 (1969).
[10] E. J. Smith *et al.*, paper CWJ1 at the Conference on Lasers and Electro-Optics, Anaheim, CA, May 10–15, 1992.

Selected Additional References

A. Laser-Induced Material Removal

S. S. Charschan, ed., *Lasers in Industry*, Van Nostrand-Reinhold, Princeton, NJ, 1972, Chapter 3.
W. W. Duley, *CO_2 Lasers: Effects and Applications*, Academic Press, New York, 1976, Chapter 4.
J. F. Ready, Effects due to Absorption of Laser Radiation, *J. Appl. Phys.* **36**, 1522 (1965).
J. F. Ready, *Effects of High-Power Laser Radiation*, Academic Press, New York, 1971.
J. F. Ready, Material Processing—an Overview, *Proc. IEEE* **70**, 533 (1982).
E. N. Sobol, *Phase Transformations and Ablation in Laser-Treated Solids*, Wiley, New York, 1995.

B. Hole Drilling

D. Belforte and M. Levitt, eds., *The Industrial Laser Annual Handbook—1990 Edition*, PennWell Books, Tulsa, OK, 1990.
S. S. Charschan, ed., *Lasers in Industry*, Van Nostrand-Reinhold, Princeton, NJ, 1972, Chapter 4.
W. W. Duley, *CO_2 Lasers, Effects and Applications*, Academic Press, New York, 1976, Chapter 5.
S. Lugomer, *Laser Technology: Laser Driven Processes*, Prentice-Hall, Englewood Cliffs, NJ, 1990, Chapter 4.
R. W. Olson and W. C. Swope, Laser Drilling with Focussed Gaussian Beams, *J. Appl. Phys.* **72**, 3686 (1992).

M. H. H. van Dijk, Laser Drilling of Gas Turbine Components, *Industrial Laser Review*, p. 7 (October 1994).
B. S. Yilbas, Study of Affecting Parameters in Laser Hole Drilling of Sheet Metals, *J. Eng. Mat. Tech.* **109**, 282 (October 1987).

C. Cutting

D. Belforte and M. Levitt, eds., *The Industrial Laser Annual Handbook—1990 Edition*, PennWell Books, Tulsa, OK, 1990.
E. Beyer and D. Petring, State of the Art in Laser Cutting with CO_2 Lasers, *Proceedings of ICALEO '90*, Laser Institute of America, Orlando, FL, 1991.
B. Biermann, S. Biermann, and H. W. Bergmann, Cutting Al-Alloys Using High Pressure Coaxial Nozzle, *J. Laser Appl.* **3**, 13 (1991).
M. Bozzo, Laser Cutting of Thin Plates, *J. Laser Appl.* **2**, 14 (1990).
G. K. Chui, Laser Cutting of Hot Glass, *Ceramic Bull.* **54**, 514 (1975).
S. Lugomer, *Laser Technology: Laser Driven Processes*, Prentice-Hall, Englewood Cliffs, NJ, 1990, Chapter 4.
L. Ohlsson *et al.*, Comparison between Abrasive Water Jet Cutting and Laser Cutting, *J. Laser Appl.* **3**, 46 (1991).

D. Scribing

S. S. Charschan, ed., *Lasers in Industry*, Van Nostrand-Reinhold, Princeton, NJ, 1972, Chapter 4.
G. Chryssolouris, *Laser Machining: Theory and Practice*, Springer-Verlag, New York, 1991, Chapter 6.

E. Marking

W. L. Arthur, Laser Marking: Rating the Players, *Lasers & Optronics*, p. 51 (October 1989).
B. H. Klimt, Review of Laser Marking and Engraving, *Lasers & Optronics*, p. 61 (September 1988).

F. Balancing

A. J. Beaulieu, Rapid Balancing of Gyroscopes with TEA CO_2 Laser, *Electro-Optical Systems Design*, p. 36 (May 1973).
H. Schneider, Balancing with Lasers, in *Industrial Applications of Lasers*, H. Koebner, ed., Wiley, New York, 1984.

H. Laser Deposition of Thin Films

E. N. Sobol, *Phase Transformations and Ablation in Laser-Treated Solids*, Wiley, New York, 1995, Chapter 6.

I. Specific Examples of Material Removal

M. Brownhill and B. Guthner, Fabricator Switches from Plasma to Laser Cutting, *Industrial Laser Review*, p. 7 (February 1995).
G. Connolly, Hydroforming/Laser Combo Cuts Component Lead Time, *MetalForming*, p. 55 (April 1994).

Chapter 17 | Lasers in Electronic Fabrication

Laser processing is widely used for fabrication of electronic components [1]. Applications include trimming of resistors, hole drilling, scribing, and marking. Lasers may also be used to fabricate microcircuits via the controlled deposition of semiconductors, metals, and insulators. This chapter will describe the status and future prospects for laser processing of electronic circuit elements.

We will emphasize currently used applications in electronic fabrication. Many of these applications involve material removal, in applications such as substrate drilling and cutting. We will also discuss developing approaches to use of lasers in integrated circuit fabrication.

A variety of different types of lasers are used in electronic fabrication. Table 17-1 summarizes the lasers that are most commonly employed. The use of the CO_2 and the infrared Nd:YAG lasers in electronic processing applications is well established; these lasers have been used for many years for applications such as trimming and drilling. The green and ultraviolet lasers may be focused to a smaller spot than the infrared devices, and they may be chosen when small focal diameter is desired. The use of ultraviolet lasers is relatively new, especially the excimer and frequency-tripled and -quadrupled Nd:YAG lasers. These lasers have become more mature and reliable, and they now present viable options for electronic processing. They offer the attractive feature of very high absorption in many materials of interest.

A. Established Applications in Electronics

Lasers have reached production status for a variety of applications in the electronics industry. One of the most significant is the trimming of resistors [2]. This can significantly increase the yield in the processing of resistive elements.

Resistors for use in electronic circuits may be printed on a substrate from a liquid ink and subsequently fired and dried. Such resistive elements are called thick film resistors. In some cases, resistors are formed by vacuum deposition of metals. These resistive elements are called thin film resistors.

Table 17-1 **Lasers for Electronic Material Processing**

Laser type	Wavelength (nm)	Operation	Typical applications
CO_2	10,600	Continuous, millisecond pulse	Hole drilling, cutting, scribing, marking
Nd:YAG	1,064	Submicrosecond pulse	Trimming, marking, package sealing
Doubled Nd:YAG	532	Submicrosecond pulse	Trimming, mask saving
Argon ion	488, 514, 364	Continuous	Deposition, lithography
Xenon ion	480–540	Microsecond pulse	Mask repair
Excimer	191, 249	Submicrosecond pulse	Deposition, lithography, material removal
Tripled and quadrupled Nd:YAG	355, 266	Submicrosecond pulse	Competitor to excimers

In either case, it is difficult to control the value of the resistance to within the tolerances required by the circuit. Thus, resistors are fabricated with intentionally low values of resistance and then trimmed by removing material from the conducting path. The value of the resistance may be monitored during the trimming process, and the trimming is terminated when the desired resistance is reached. Alternatively, the performance of the circuit may be measured during trimming; in this case, the trimming is terminated when the circuit is operating properly.

Trimming by laser ablation of material from the resistors has become widespread. The technique most often uses a repetitively Q-switched Nd:YAG laser, emitting a train of pulses with peak power in the kilowatt range, pulse durations around 250 nsec, and pulse repetition rates in the kilohertz regime. Typical parameters for the trimming of thin and thick film resistors with Nd:YAG lasers are presented in Table 17-2. The Nd:YAG laser offers the advantage of smaller cut widths than the carbon dioxide laser, which has also been used. The infrared Nd:YAG laser is perhaps the most common choice, but if fine trimming cuts are desired, the frequency-doubled Nd:YAG laser is employed. Trimming with excimer lasers has also been demonstrated, but this has not yet become common.

Various geometries for cutting have been used. A popular geometry is the L-shaped cut. The resistor is cut partway across, and then the direction of travel is changed by 90°. This leads to finer control of the process. If one simply continued to cut straight across the resistor, there would be relatively poor control of the tolerance when the cut reached almost all the way across the resistor. The percentage change in resistance per unit cut length would become very large. The L-shaped cut allows a more gradual approach to the final desired value.

Table 17-2 **Thin and Thick Film Resistor Trimming**

Typical Values for Nd:YAG Laser Trimming of Thin Film Resistors	
Peak power	1200 W
Repetition rate	10 kHz
Pulse duration	250 nsec
Cut width	0.001 in.
Pulse overlap	65 percent
Typical Values for Nd:YAG Laser Trimming of Thick Film Resistors	
Peak power	4000 W
Pulse repetition rate	3 kHz
Pulse duration	250 nsec
Cut width	0.002 in.
Overlap	95 percent

Thermal shock and damage along the cut, which cause drift of the resistor after trimming, have been problems. Keeping the laser power down to the minimum needed for a clean cut will reduce problems of damage and post-trim drift. With careful control, the change in resistance after the trimming can be in the range around 0.1 percent or less.

Laser trimming of resistors is especially notable because it is a process that has become routine in the electronics industry, and it has essentially replaced the older method, which was abrasive trimming. Laser trimming offers the advantages of better cleanliness and better control over the final resistance. The laser trimming operation results in a higher yield; that is, a larger fraction of the resistors are within the prescribed tolerance.

Laser-based removal of material from substrates is also widely used for shaping, drilling, and cutting of substrates. We have already discussed these applications involving material removal in Chapter 16, but we will review them here with emphasis on application in fabrication of electronic components.

Pulsed carbon dioxide lasers are commonly used to drill holes in alumina substrates for circuit boards. Alumina ceramics absorb well at the CO_2 laser wavelength. Laser drilling in hard, high-temperature-fired alumina ceramic is attractive because drilling ceramics by conventional means is expensive. It usually requires diamond-tipped drill bits, which are subject to breakage. Conventional drilling of holes less than 0.25 mm in diameter is difficult. Laser drilling easily produces small holes in this brittle material without danger of fracturing it. When a pulsed CO_2 laser is used with alumina of thickness less than 1 mm, typical hole diameters may be in the range of 0.1 to 0.2 mm with tolerances around 0.02 mm.

Because of the brittleness of these ceramics, conventional hole drilling is usually performed on the material before it is fired. When it is fired, the dimensions may change. Laser drilling is performed after firing, thus eliminating this potential problem.

Automated laser systems drill complicated patterns of holes in material about 0.6 mm thick. The system includes fixturing to hold the ceramic, a stepping motor to move the fixture to predetermined positions at which the holes are to be drilled, and a control to turn the laser on and off so that holes are drilled in the prescribed positions. The entire process may be automated to yield as a final result the desired pattern of holes in the ceramic. Figure 17-1 shows examples of laser-drilled holes in alumina substrates.

Lasers are also used to cut or shape a wide variety of nonmetallic substrate materials. Many cutting operations have employed carbon dioxide lasers because of their capability for producing high values of continuous power and because the CO_2 laser radiation is absorbed well by ceramics. Cutting of nonmetallic materials, such as alumina or silicon, proceeds easily. A continuous CO_2 laser operating at a level around 100 W is adequate for many cutting applications. The advantages of laser cutting include lack of tool wear, reduced loss of material around the cut, higher yield because of reduced breakage and rapid turnaround because of the ease of fixturing.

An alternative to cutting of materials like ceramic is scribing. In scribing, a series of closely spaced or overlapping blind holes is drilled partway through the sample. The material will then easily snap along the path defined by the laser scribe line. Usually, scribing is done in a straight line in order to provide a clean separation, without random cracking, in the desired path.

Scribing is widely used for applications like dicing alumina substrates into chip carriers or separating silicon wafers into chips. The scribing of alumina uses carbon

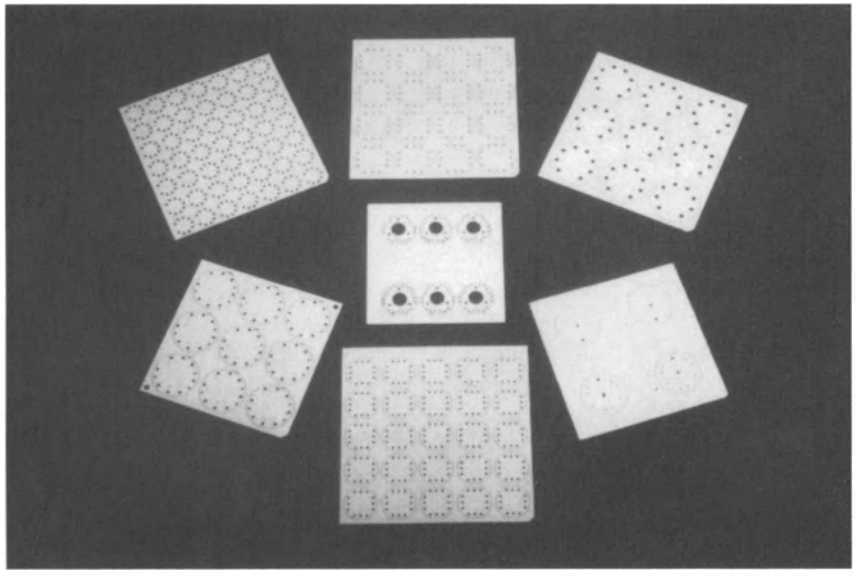


Figure 17-1 Examples of patterns of laser-drilled holes in alumina ceramic substrates. (Photograph courtesy of Electro Scientific Industries, Inc.)

dioxide lasers and the scribing of silicon is done with Nd:YAG lasers, because the absorption of the different materials is high at the different wavelengths. The motivation for scribing rather than cutting is that the scribing rate is higher than the cutting rate. For alumina around 0.025 in. thick, the material may be scribed at rates around 10 in./sec with an intermediate-power CO_2 laser, whereas cutting rates may be a few tenths of an inch per second for a similar laser. Scribing also offers the advantage that a substrate may be scribed before processing is completed and then easily separated into chips after processing. Table 17-3 presents typical scribing parameters for alumina ceramic and silicon.

There are many requirements for identification of manufactured parts in electronic fabrication. Product marking is carried out for purposes of identification or imprinting product information. Conventional techniques include printing and stamping. Laser marking provides an attractive alternative for durable permanent marks with no possibility of contamination.

Two different laser-based approaches to marking of electronic products are employed. In the first method, one scans the focused high-power beam over the surface of the part and modulates it to vaporize a small amount of material at selected positions. The mark thus consists of a dot pattern. In the second method, called image micromachining, the pattern is formed all at once, by delivery of a high power beam already focused in the desired pattern. The first method is applicable to marking of materials that require relatively high energy to vaporize, for example, silicon wafers. The second method is applicable for softer materials that require lower energy per unit area, such as packaging materials.

Laser printing of identifying numbers on silicon wafers is attractive because of the brittleness of the material, because subsequent manufacturing operations could

Table 17-3 **Parameters for Laser Scribing**

Typical Values for CO_2 Laser Scribing of Alumina	
Peak power	300 W
Pulse repetition rate	1 kHz
Pulse duration	100 μsec
Hole diameter	0.005–0.006 in.
Hole depth	0.006–0.007 in.
Scribe speed	6 in./sec
Typical Values for Nd:YAG Laser Scribing of Silicon	
Peak power	500 W
Pulse repetition rate	3.5 kHz
Pulse duration	300 nsec
Scribe speed	6–8 in./sec
Overlap	80 percent

destroy printed marks, and because marking with inks represents a potential source of contamination. Nd:YAG lasers are generally used for this application because silicon absorbs well at wavelengths near 1 μm, whereas it does not have high absorption at the carbon dioxide laser wavelength.

Laser vaporization is also used to remove small amounts of unwanted metallization, for example, on master patterns used for making photographic masks for integrated circuit chips. This use, which has reached routine production status, is employed to reduce the number of defects per mask, and thereby to increase the yield of the process. One desires typically a short-wavelength laser, often a frequency-doubled Nd:YAG laser operating at 532 nm. The absorptivity of the metal is relatively high at this wavelength. Because of the short wavelength, the beam may be focused to a very fine spot.

The masks are subjected to visual inspection under a microscope with a display on a television monitor. The surface of the mask is scanned in a pattern. When a spot of undesired metallization is located, it is lined up in the center of the display with grid marks to indicate the aiming point. The diameter of the focused beam is adjusted to accommodate the size of the particular spot of metallization. The laser is then fired, delivering a pulse that vaporizes the unwanted metallization.

In addition to drilling holes in substrate materials, laser ablation is used to fabricate vias for the fabrication of multichip modules and in printed wiring boards. The vias are then metallized in order to form an electrical contact from one side of the material to the other.

In the production of multichip modules, vias are formed in polyimide by excimer laser ablation. Excimer lasers have been used for this application because of the very high absorption of plastic materials in the ultraviolet portion of the spectrum. Typical laser parameters used have been 150–300 mJ/cm^2 at 308 nm wavelength (XeCl excimer laser). Projection through a mask has allowed drilling of many holes at the same time on fields in excess of 100 mm square. This technique has been used to fabricate many billions of 6-μm-diameter vertical interconnection vias. The laser drilling has proved to be faster and less expensive than the competing techniques of wet etching or reactive ion etching.

Laser drilling has also been investigated for drilling printed wiring boards. In such boards, the density of conductors is limited by the presence of relatively large pads. Pad size is determined by the diameter of vias under the pads. Reduction of via and pad size is desired to increase component density.

Current technology has used numerically controlled mechanical drillers. Costs have been high, and they increase rapidly as the via size decreases. The minimum practical via size is about 150 μm with mechanical drilling.

Several laser types have been investigated for this application. Excimer lasers have excellent absorption in polymeric materials. They can produce very clean holes in the 6–60 μm diameter range. Removal rates are low, so that costs are high except in fixed designs with mask imaging where many vias can be drilled simultaneously. This application has reached limited production.

Frequency-tripled and -quadrupled Nd:YAG lasers can drill vias with diameters in the 25–50 μm range. Drilling rates are faster and costs lower than for excimer lasers, but more precise control is needed. These lasers appear to be competitive for via drilling.

Carbon dioxide lasers are also candidates for via drilling. The absorption is high at the CO_2 laser wavelength, and the drilling rate is high. The hole sizes are larger than for the shorter-wavelength lasers, but CO_2 lasers appear economically competitive.

Another application of laser ablation is the cutting of links in a circuit. In circuits like arrays of memory cells, one may insert extra columns or rows of elements in the array. The devices are tested, and faulty devices are identified. Then, the laser is used to vaporize the conductors that link the rows containing faulty elements. The faulty devices are thus removed from the circuit. This technique has been widely used to repair memory arrays. It provides redundancy for dynamic random access memory (DRAM) yield enhancement. It has also been used for personalization of multichip modules and for application specific integrated circuits (ASICs). Frequency-doubled Nd:YAG lasers are frequently used for such repair and personalization.

Lasers have also been used for link formation. Multilayer substrates with an incomplete top layer are fabricated. The top layer may be only bond pads. Lower layers are generic, design-independent wiring patterns. Spot links may be fabricated in a maskless deposition process. This procedure is useful for applications like customization through simple laser spot links, fast turnaround processing, fabrication of ASICs, and personalization of multichip modules. Frequency-doubled Nd:YAG lasers are also used in this application.

One approach to laser link making has been to use materials like copper formate. A solution of copper formate in water is spun onto a wafer, and a solid film is dried from the solution. The film is decomposed in selected areas by a frequency-doubled Nd:YAG laser to leave copper metallization, which forms the interconnecting links. The remaining film is then washed off with water. Other metal-organic compounds are also available if other types of metal interconnects are desired.

In another interconnection application, laser tape-automated bonding (TAB) of leads to integrated circuit input–output pads offers advantages compared with conventional techniques like thermocompression bonding. Because of its localized nature, laser bonding has less chance of damaging the chip. It has also provides stronger bonds. TAB has commonly been performed with millisecond pulse Nd:YAG lasers. Materials that have been used include gold, copper, and tin. The rate of formation of the bonds can be high, greater than 60 per second. This application is still relatively small but appears to be growing.

Another application involving material removal is the stripping of insulation from wires, in preparation for making an electrical contact to the end of the wire. Laser ablation of the insulation on wires can be performed in a wide variety of cut patterns and can remove hard or high-temperature insulation. It does not cause nicks, scratches, or broken wire strands, as mechanical removal of insulation sometimes does. Laser stripping of wires has reached production status. It has most frequently

been performed with CO_2 lasers, but recent studies have indicated that excimer lasers can produce clean, well-defined cut patterns, with very little residue. For stripping of polyurethane insulation in particular, the use of excimer lasers can provide greater precision and cleanliness than CO_2 lasers.

A final packaging application involves hermetic sealing of electronic packages by laser welding. This has become a common production application, especially for the sealing of microwave packages. Previously, tungsten-inert gas (TIG) welding and resistance welding were common, but these techniques produce large heat-affected zones and deliver relatively large amounts of heat to the package. The advantage of laser welding is the very localized heating involved, so that there is no damage or distortion of the delicate components in the package. The heat-affected zone associated with laser welding is smaller than for the competing techniques. Laser welding provides production package sealing with excellent hermeticity and good shock and vibration tolerance. The leak rates may be as low as 10^{-9} atm cc He/sec, a value that satisfies almost all hermeticity requirements.

One common material for microwave packages is 6061 aluminum. This alloy is somewhat brittle, so that use of a filler material is desirable to avoid cracking of the weld. Aluminum alloy 4047, containing about 12 percent silicon, has proved useful as a filler.

Packages are often welded by an infrared Nd:YAG laser with pulses in the millisecond regime. Typical welding conditions for aluminum packages might be an average power of a few hundred watts, pulse duration of a few milliseconds, and pulse repetition rates of a few tens of hertz. Seam rates might them be in the 50–100 in./min range. Hermetic sealing of packages in production by laser welding has become common practice, especially when components in the package are heat sensitive.

The foregoing examples present only a few of the many practical applications of laser processing that have become established for fabrication of electronic structures and components.

B. Applications in Integrated Circuit Fabrication

Laser processing has been used in the electronics industry for many years for applications such as described in the previous section. The applications discussed to this point have involved components in an electronic assembly, such as resistors and substrates, packaging of the devices, and repair. We have not yet discussed applications involving fabrication of integrated circuits. Fabrication of integrated circuits has been performed by optical lithography for decades. In the photolithography process, photoresist is exposed through a mask that is imaged on the surface of the resist to produce the features of the microcircuit. Current sources for photolithography employ mercury lamps to a large extent.

Feature size in integrated circuits has steadily decreased, and it will continue to decrease, with concomitant increase in chip density. Significant advances in optical lithography technology have been required to make this possible. The shrinking fea-

ture sizes in microelectronics require that lithographic sources move to shorter wavelengths, to the deep ultraviolet. Feature sizes in the mid 1990s are in the 0.35 μm regime. But conventional light sources (mercury lamps) cannot deliver enough energy with the required resolution, as feature sizes continue to shrink, to the 0.25 μm region and below. In addition, the relatively broad linewidth of the mercury lines is a problem. There are very few optical materials suitable for use at wavelengths of 250 nm or less. These include fused quartz and a few fluoride materials (calcium fluoride, lithium fluoride, etc.). Only one material, fused quartz, is suitable for use in high-quality optical projection systems. With only one material available, one cannot design an optical system that incorporates color correction. Thus, the linewidth of the optical source must be very narrow.

The current mercury lamps, operating at a wavelength of 365 nm, should be adequate to support the production of 64-megabit dynamic random access memory (DRAM) chips with feature sizes of 350 nm, which are being developed in the mid 1990s, but higher-density chips will require a feature size of 250 nm or less. The DRAM, which may be regarded as a benchmark of microcircuit fabrication, has been quadrupling in density every three to four years. By the early 2000s, we may anticipate the need for technology to fabricate 1-gigabit DRAMs. New technology will be required to meet this need.

Deep ultraviolet laser sources are moving into photolithographic applications for high-density microcircuitry. There is a major thrust toward laser-based photolithography in the deep ultraviolet portion of the spectrum. Work so far has emphasized use of the KrF excimer laser at 248 nm.

Laser sources offer desirable features for deep ultraviolet photolithography. Their high power output leads to short wafer exposure times. The linewidth may be as narrow as a 1–2 picometers (pm) for KrF lasers that are frequency stabilized by means of an element like an internal grating. This meets the linewidth requirements for the source, so that it is possible to design high-performance all-quartz optical systems without the need for color correction.

The use of KrF-laser-based photolithography systems has been studied at several institutions. The results indicate that such systems can meet the requirements of the industry for fabrication of devices with feature size of 0.25 μm, with adequate control of critical dimensions, overlay, and depth of focus [3]. Thus, development of 256-megabit DRAMs will probably be carried out using KrF-laser-based photolithography.

Some experimental work has also begun on ArF excimer laser photolithography, at a wavelength of 193 nm. This would reduce the possible feature size still further. Projections have suggested that optical lithography may reach an ultimate limit at a feature size around 120 nm, perhaps around the end of the first decade of the twenty-first century. At that time a different technology, perhaps electron beam lithography, would be required to reduce feature size further. But it appears likely that excimer-laser-based microlithography will play an important role in the fabrication of microcircuits from the late 1990s until around 2010.

Another important research area in recent years has been the investigation of laser processing of semiconductor circuitry at the wafer level. This work, too, offers

the possibility of extending the use of lasers to fabrication of the integrated circuit itself. The development of low-temperature processing technology is a key requirement. Laser radiation can provide fast localized heating that does not affect nearby portions of the circuitry.

Figure 17-2 illustrates some of the possible steps where lasers can play a roll in the fabrication of semiconductor integrated circuits. The figure shows some representative processing steps common in the manufacture of semiconductor circuits but does not represent fabrication of any specific circuit. The figure also illustrates some of the ways in which laser processing could applied to these steps. In the following discussion, we will emphasize two of these in detail: the annealing of ion implantation damage, and laser photodeposition of circuit patterns.

The ideas for direct laser intervention in the fabrication of integrated circuits began with research into annealing of damage produced by ion implantation in semiconductor circuits [4–7]. Ion implantation has become the preferred method for introducing impurity atoms into silicon to define circuit structures. The implantation performed with high-energy (perhaps 200 keV) ions damages the crystalline lattice of the silicon and makes it amorphous near the surface. Annealing is necessary to recrystallize the silicon. The annealing is conventionally done in a furnace. Laser annealing can also recrystallize the silicon and would have the advantage that the entire wafer is not heated. Laser annealing of ion implantation damage has been studied very extensively. It was found that laser annealing can regrow the crystalline structure, producing single-crystal silicon. Laser annealing removes defects, such as dislocation loops, which are common in furnace regrown silicon. It activates the implanted ions and thereby lowers the surface resistance. In addition, in contrast to furnace annealing, it can be done in such a way as to preserve the profile of the implanted ions.

Laser annealing can be accomplished either by applying a very short (100 nsec) pulse of energy over an area of several square centimeters, or by scanning a continuous beam of small area (perhaps 10^{-4} cm^2) over the surface of the material being processed. In the continuous case, the duration of processing for a particular area may be of the order 10^{-3} sec. This is much longer than in the pulsed case, but still short enough to retain the intrinsic advantages in comparison with conventional methods. Pulsed laser annealing of ion implanted layers has most often used Q-switched ruby or Nd:YAG lasers with pulse durations around 100 nsec. Continuous wave laser annealing is performed with a beam from a continuous laser, most often an argon laser. The beam is scanned over the surface of the sample so as to provide continuous coverage. The laser power and scan speed are controlled to produce surface heating without melting. Thus, the process involves solid state regrowth, similar to the process that occurs in furnace annealing; for laser annealing, though, the heating is confined to a very thin surface layer, and the time scale is much shorter.

Laser annealing of ion implantation damage with both the continuous method and the pulse method activates implanted ions and regrows the crystalline silicon well. In addition, the continuous laser annealing preserves the profile of the implanted ions. Figure 17-3 shows the distribution of 50 keV boron ions implanted in

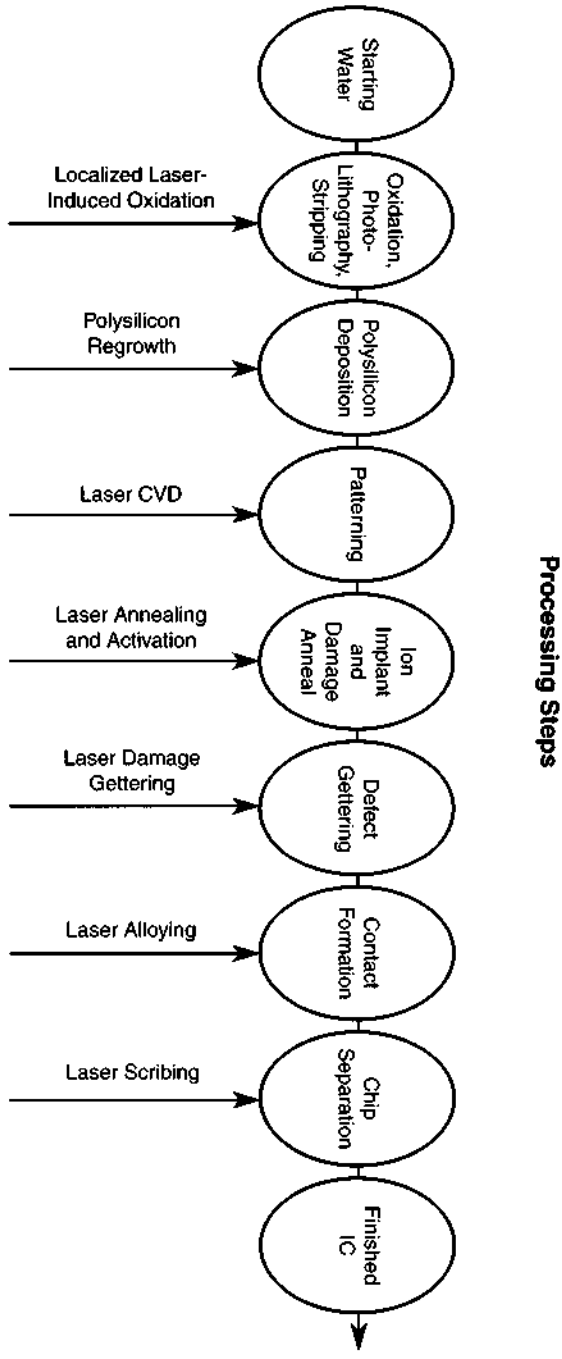


Figure 17-2 Typical processing steps for microcircuit fabrication and possible areas of laser impact.

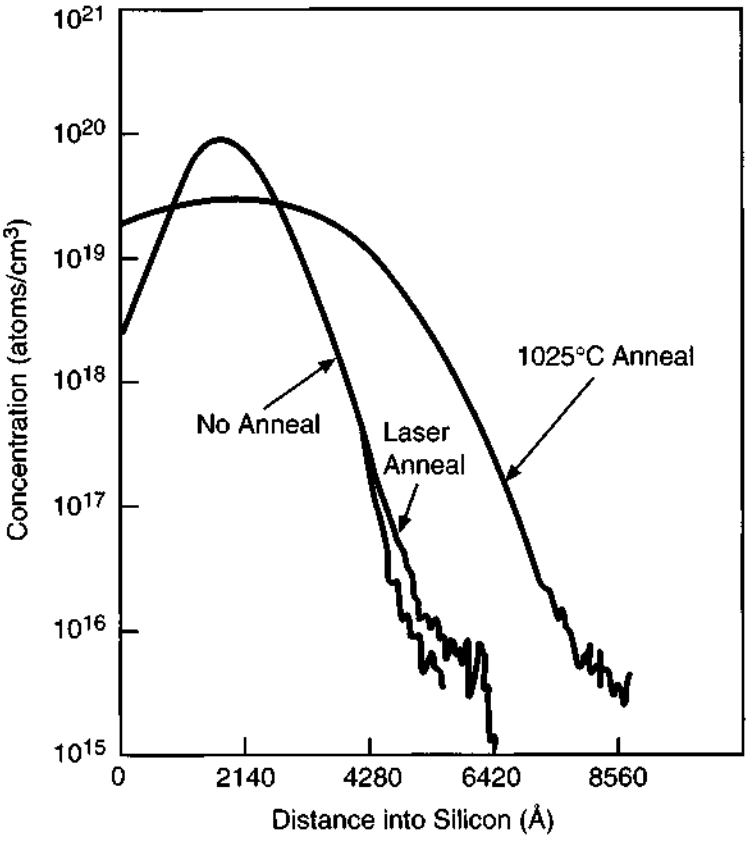


Figure 17-3 Distribution of implanted 50 keV boron ions in silicon, as implanted, after laser annealing, and after 1025°C furnace annealing. (From J. F. Ready, B. T. McClure, and W. L. Larson, *Scientific Honeyweller* **2**, 37 (September 1981).)

silicon. The original distribution of the spatial profile of the implanted boron ions is indicated. After laser annealing, the distribution is changed insignificantly. However, after an anneal at 1025°C in a furnace, the profile has changed substantially, as indicated. This means that the ions have migrated through the silicon. This will have the effect of broadening the feature sizes and will be incompatible with circuit structures having extremely small feature sizes.

Laser annealing of ion implantation damage in semiconductors was a popular research topic in the late 1970s and early 1980s. However, it has not yet become widely used in the microelectronic industry because of economic factors. Laser annealing of ion implantation damage has remained more expensive than furnace annealing.

Another potential application of laser technology in integrated circuit fabrication is the direct generation of patterns. For example, laser-aided chemical vapor deposition, which can be used to deposit layers of semiconductors, metals, and insulators,

can define the circuit features. The deposits can have dimensions in the micrometer regime, and they can be produced in specific patterns. Laser chemical vapor deposition can employ two approaches:

- A pyrolytic approach in which the heat delivered by the laser breaks chemical bonds in vapor phase reactants above the surface, allowing deposition of the reaction products only in the small heated area.
- Direct photolytic breakup of a vapor phase reactant. This approach requires a laser with proper wavelength to initiate the photochemical reaction. Often, ultraviolet excimer lasers have been used. One example is the breakup of trimethyl aluminum gas with an ultraviolet laser to produce free aluminum, which deposits on the surface. Again, the deposition is only on the localized area that the beam strikes.

A typical example might involve use of a krypton fluoride excimer laser operating at 249 nm with a pulse duration around 100 nsec and a pulse repetition rate that can be varied up to 200 Hz. For metal deposition, a fluence in the range from 0.1 to 1 joule per square centimeter per pulse is typical.

The gas handling system is usually a dynamic flowing system, because the reactants in the gas chamber are constantly being replenished and by-products are removed. In contrast, in a static system, reactants are depleted and by-products build up.

A variety of reactions have been investigated [8–11]. Table 17-4 shows a representative sample of such reactions. This list is not exhaustive by any means, but it does show the typical types of processing that can be performed. It is apparent the all the operational steps required to fabricate an integrated circuit are available. One can deposit materials, including metals, insulators, and semiconductors. One can perform doping operations, etching, and stripping of insulating materials. All these operations can be done in a highly controllable fashion, depositing a material only on the areas where it is desired.

Figure 17-4 shows a schematic diagram for a computer-controlled system to deposit the entire integrated circuit on a wafer. When it is desired to deposit metallization, for example, the chamber is pumped down and the gas valve is opened to allow a suitable metal-organic gas, such as trimethyl aluminum, into the system. The beam is scanned over the surface of the wafer and modulated on and off so as to deposit metallization on all the areas where it is desired. When the metallization is complete, the system may be pumped down again, a new gas allowed in, and new patterns of some other material formed. This technique essentially would allow one to direct write a circuit onto the substrate [12,13].

Compared with conventional processing techniques, laser deposition allows fabrication of entire circuits all in a clean vacuum system without a need to move back and forth between different processing stages. Thus, it is a very clean technology capable of producing circuitry with very low levels of defects.

The development of laser technology for direct generation of integrated circuits would provide an effective method of cool dry processing for microelectronics, useful in the preparation of solid state devices and their packaging.

Table 17-4 **Laser Photochemical Processes**

Process	Reaction
Metallization	
Al	$2Al(CH_3)_3 + 3H_2 \xrightarrow{h\nu} 2Al + 6CH_4\uparrow$
W	$WF_6 + 3H_2 \xrightarrow{h\nu} W + 6HF\uparrow$
Insulator Deposition	
SiO_2	$SiH_4 + 2N_2O \xrightarrow{h\nu} SiO_2 + 2N_2\uparrow + 2H_2\uparrow$
Si_3N_4	$3SiH_4 + 4NH_3 \xrightarrow{h\nu} Si_3N_4 + 12H_2\uparrow$
Semiconductor Deposition	
Si	$SiH_4 \xrightarrow{h\nu} Si + 2H_2\uparrow$
Si	$Si_2H_6 \xrightarrow{h\nu} 2Si + 3H_2\uparrow$
GaAs	$Ga(CH_3)_3 + AsH_3 \xrightarrow{h\nu} GaAs + 3CH_4\uparrow$
Doping	$2PH_3 + 2Si \xrightarrow{h\nu} 2Si{:}P + 3H_2\uparrow$
Etching	$Si + 4HCl \xrightarrow{h\nu} SiCl_4\uparrow + 2H_2\uparrow$
	$Si + 2Cl_2 \xrightarrow{h\nu} SiCl_4\uparrow$
Stripping	$SiO_2 + Si \xrightarrow{h\nu} 2SiO\uparrow$

C. Summary

The current status of laser applications in microelectronic fabrication is that laser technology has made substantial impacts on the fabrication of components such as the substrates and resistors. It is widely used for trimming both thick and thin film resistors, for scribing wafers, for hole drilling in substrates, for welding of hermetically sealed packages, and for stripping insulation from wires. The marking of silicon wafers with identification numbers has also become well established. In all these applications, lasers have become established production tools, replacing earlier technology for many applications.

The applications of lasers in microelectronic fabrication have become accepted on a routine production basis because they offer an economic advantages in many cases. Laser fabrication is an effective economic competitor for applications such as resistor trimming.

In the areas of laser annealing and direct generation of patterns, the economic factors do not yet favor laser processing. At present, laser annealing of ion implantation is more expensive than furnace annealing. Hence, it has not been widely adopted except in a few specialized niche areas. It also faces strong competition from other regrowth techniques, including regrowth using strip heaters and arc

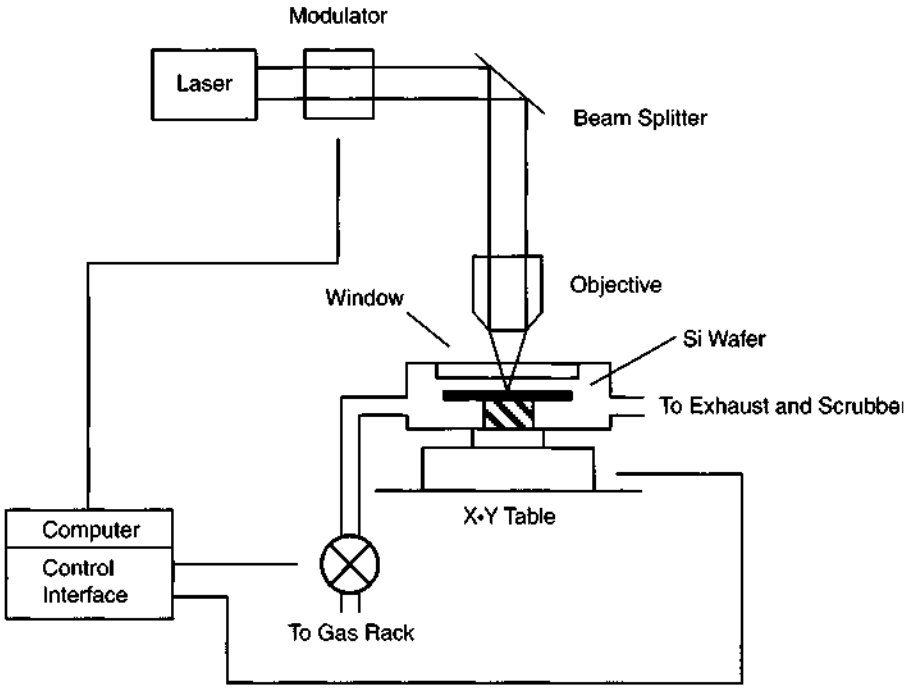


Figure 17-4 A computer-controlled system for direct deposition of integrated circuits.

lamps. These competing techniques have also yielded very favorable results. For the future, as feature sizes in integrated circuits continue to shrink, laser annealing may be needed to minimize the motion and diffusion of the implanted ions so as to reach the lowest possible values of feature size.

Similarly, laser deposition of the integrated circuits on an entire wafer has been relatively slow and, hence, expensive. In addition, the process equipment used in the semiconductor industry represents a large investment that the manufacturers are reluctant to make obsolete. Therefore, laser direct write processing of integrated circuits has not reached production status. Near term applications will remain specialized; we may look for uses such as customization and personalization of integrated circuits. In these applications, selected pathways of metallization may be deposited or removed so as to change the operation of a specific circuit. Also, laser repair of expensive circuits is developing. One can replace missing metallization, for example, and hence salvage a high-value circuit that would be otherwise scrapped.

For the longer term, we look toward advances in laser technology to provide high-power lasers that can be used with projection techniques to fabricate the entire circuit through a mask. This will increase the throughput and should remove existing restrictions on the time that it takes to fabricate an entire integrated circuit by direct write processing. With such advances, it is likely that laser deposition can become an effective economic competitor.

In conclusion, laser processing has provided many cost-effective solutions to problems in microelectronic fabrication. We may expect these applications to burgeon in the future as laser processing plays more of a role in fabrication of the integrated circuit elements themselves.

D. A Specific Example: Laser-Based Photomask Repair

The use of lasers to repair photomasks has become widespread in the microelectronics industry. Photomasks used in integrated circuit production are typically patterns of chromium metallization on clear glass substrates. Such masks are subject to various defects, including excess chromium metallization, either as isolated spots or as extensions of metal from a feature, and areas of missing metallization. The ability to repair mask defects offers an important cost saving alternative to discarding the defective mask. It results in higher productivity and faster turnaround in the mask making process.

A mask saving system that can repair defects involving either excess or missing metallization is shown in Figure 17-5. The mask is scanned microscopically, and the

Figure 17-5 Production system for location, identification, and repair of photomask defects. The system can both remove excess metallization and fill in missing metallization. (Photograph courtesy of Quantronix Corp.)

image of the mask features is displayed on a television monitor. The operator can locate and identify the mask defects, and then repair them.

Areas with excess chromium metal are repaired by vaporization of the metal with a frequency-doubled Q-switched Nd:YAG laser beam. It is possible to remove excess chromium within feature spacings less than 1 μm.

Areas lacking metallization are repaired by a pyrolytic deposition process. A gas mixture containing chromium and molybdenum hexacarbonyls enters the system, and it is decomposed by the heat from an acoustooptically modulated continuous argon laser beam, which is scanned over the area of missing metallization. This results in the deposition of a metallic film that fills in the area. An example is shown in Figure 17-6. The left portion of the figure has a missing 10 μm square area of metallization. In the right portion of the figure, the missing metallization has been filled in through the laser deposition process. The laser-deposited films are suitable for use in the photolithographic process.

The repair of masks in this way can reduce waste and increase profitability in integrated circuit fabrication.

References

[1] S. S. Charschan, ed., *Lasers in Industry*, Van Nostrand, New York, 1972.
[2] M. Oakes, *Opt. Eng.* **17**, 217 (1978).
[3] G. B. Elder, L. C. Litt, and J. G. Maltabes, *Solid State Tech.*, p. 51 (June 1995).
[4] S. D. Ferris, H. J. Leamy, and J. M. Poate, eds., *Laser–Solid Interactions and Laser Processing*, American Institute of Physics, New York, 1979.

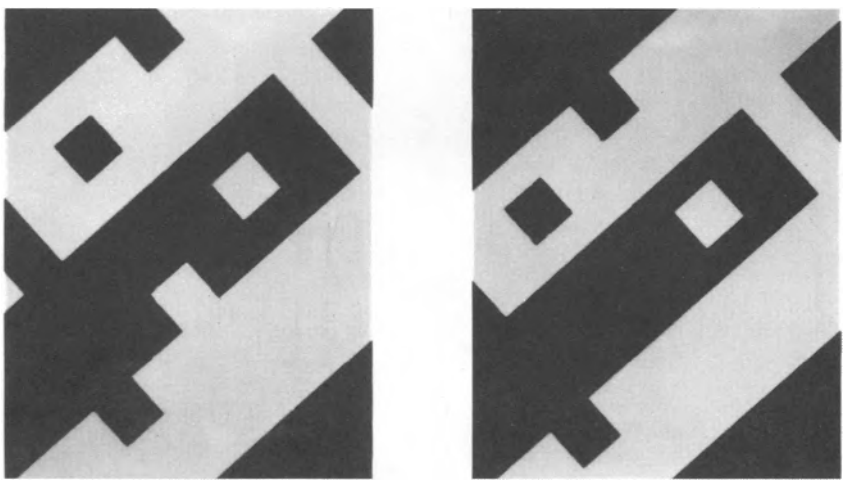


Figure 17-6 Repair of missing metallization by laser deposition. The left photograph shows a 10 μm square area of missing chromium. The right photograph shows the area filled in by laser deposition. (Photographs courtesy of Quantronix Corp.)

[5] C. W. White and P. S. Peercy, eds., *Laser and Electron Beam Processing of Materials*, Academic Press, New York, 1980.
[6] J. F. Gibbons, L .D. Hess, and J. W. Sigmon, eds., *Laser and Electron Beam Solid Interactions and Materials Processing*, Elsevier Science, Inc., New York, 1981.
[7] J. M. Poate and J. W. Mayer, eds., *Laser Annealing of Semiconductors*, Academic Press, New York, 1982.
[8] D. Bäuerle, *Chemical Processing with Lasers*, Springer-Verlag, Berlin, 1986.
[9] A. W. Johnson, D. J. Ehrlich, and H. R. Schlossberg, eds., *Laser Controlled Chemical Processing of Surfaces*, Elsevier Science, Inc., New York, 1984.
[10] R. M. Osgood, S. R. J. Brueck, and H. R. Schlossberg, eds., *Laser Diagnostics and Photochemical Processing for Semiconductor Devices*, Elsevier Science, Inc., New York, 1983.
[11] I. W. Boyd, *Laser Processing of Thin Films and Microstructures*, Springer-Verlag, Berlin, 1987.
[12] D. J. Ehrlich, *Solid State Tech.* **28** (12), 81 (December 1985).
[13] B. M. Williams *et al.*, *Appl. Phys. Lett.*, **43**, 946 (1983).

Selected Additional References

B. Braren, ed., *Lasers in Microelectronic Manufacturing*, SPIE Proc., Vol. 1598, SPIE, Bellingham, WA, 1991.
T. A. Brunner, ed., *Optical/Laser Microlithography VIII*, SPIE Proc., Vol. 2440, SPIE, Bellingham, WA, 1995.
S. S. Charschan, ed., *Guide to Laser Materials Processing*, Laser Institute of America, Orlando, FL, 1993, Chapter 8.
D. J. Ehrlich and J. Y. Tsao, *Laser Microfabrication, Thin Film Processes and Lithography*, Academic Press, Boston, 1989.
K. Jain, *Excimer Laser Lithography*, SPIE, Bellingham, WA, 1990.
A. Kestenbaum, *Semiconductor Processing Applications with Lasers*, Electrochemical Society, New York, 1979.

Chapter 18 | Principles of Holography

Holography is a method of lensless three-dimensional photography that was first invented in 1948. It was, however, only with the development of the laser in the early 1960s that holography really became practical; the bright, coherent, monochromatic laser was needed before truly good holograms could be made.

A. Formation of Holograms

In ordinary photography, one records only the amplitude of the light wave; in holography, one preserves both the amplitude and phase of the light. This implies that a hologram has preserved more of the information carried by the light. As a result, a hologram will produce a true three-dimensional representation of the light wave, just as it would have come from the original object.

Holography involves a two-step process:

1. Recording of a complex interference pattern that is produced by the superposition of two light waves. The interference pattern is recorded in a suitable recording medium, usually photographic film.
2. Reconstruction of an image from the interference pattern. This is done by illuminating the recorded pattern with a light wave, identical to one of the light waves used in the original recording process. The result of the reconstruction is an image that is a duplicate of the original object. It includes the full depth and perspective of the three-dimensional object.

The recording process begins by combining a coherent monochromatic light wave with another beam that comes from the object, usually reflected by the object. The arrangement will often be similar to what is shown in Figure 18-1. The monochromatic coherent beam from the laser is expanded by a collimator. It is then divided into two parts by a beam splitter (a partially transmitting and partially reflecting surface). One of the resulting beams is sent directly to the photographic film by reflection from a mirror. This beam is referred to as the reference beam.

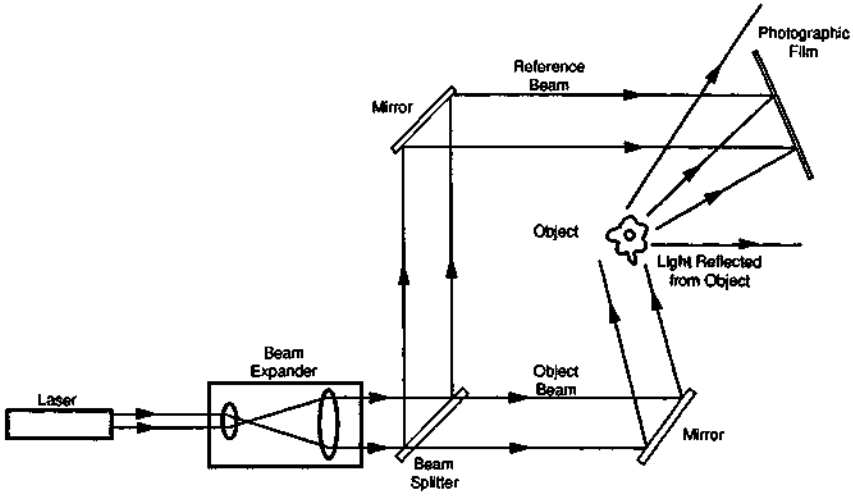


Figure 18-1 Typical arrangement for recording a hologram.

The other beam strikes the object and is reflected from it. This reflected light carries information concerning the object. The reflected light waves contain a pattern that is characteristic of the object. If one were to view this light with the eye, one would see the object. Some of this reflected light will reach the photographic film. This light is called the object beam. The light waves from both the object beam and the reference beam are superimposed at the location of the film. There, they form a complicated interference pattern, which consists of a multitude of fringes produced by constructive and destructive interference of the two beams. The shape and intensity of the fringes depend on the phase and amplitude of the two different waves. Because one of the beams carried information about the object, the interference pattern also contains information about the object.

The photographic film is exposed to the light and then developed with conventional techniques. The film preserves the interference pattern produced by the two beams and, hence, preserves the information about the object. The resulting piece of developed film is the hologram. It appears like a fogged negative; visual inspection reveals nothing that looks like the object.

There are many possible arrangements of the apparatus for making a hologram. Holograms may be made in a wide variety of configurations, with parallel, diverging, or converging beams. Different names are applied to different arrangements. However, probably the most common arrangement is similar to that shown in Figure 18-1. This is called off-axis holography, because the reference and object beams arrive at the film traveling at a nonzero angle to each other. Also, although holography strictly does not require the use of lenses, they often are used for convenience to expand the laser beam to cover a larger area, as the figure illustrates.

A common arrangement for reconstructing (or viewing) the hologram is shown in Figure 18-2. The hologram (the developed film) is reilluminated by the reference

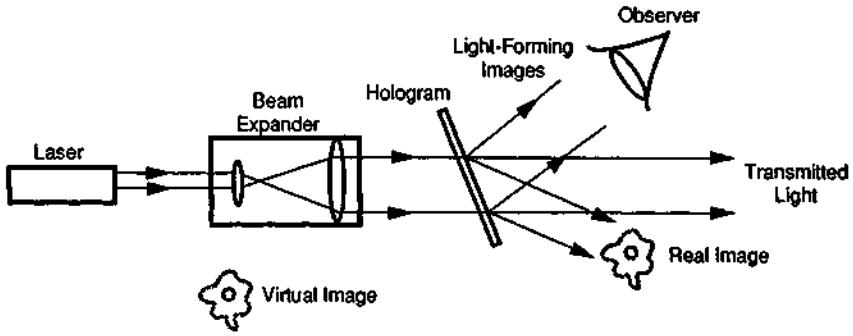


Figure 18-2 Typical arrangement for reconstructing a hologram.

beam alone. One may remove the object, or block the object beam, or remove the hologram completely to a different location and illuminate the hologram with a laser beam traveling at the same angle to the film as the original reference beam. This beam is called the reconstructing beam.

When the hologram is illuminated by the laser light at the same angle as the original reference beam, two images of the object are produced by light diffracted by the hologram. One may consider the original process of hologram formation as a method of producing a type of diffraction pattern, a closely spaced array of interference fringes, the exact shape of which depends on the nature of the object. When this diffraction pattern is illuminated, the diffracted light has properties that are dictated by the original object. Later, we will describe in more detail how the process occurs and show that the diffracted light will in fact reproduce the light waves that came from the object.

Let us consider the two different images of the object that are formed during reconstruction of the hologram. An observer will see an undistorted view of the object, just as if it were still present. The image is three-dimensional; that is, one may look partway around the object. This image is called a virtual image because it requires a lens (the lens of the observer's eye) to form it. According to elementary optics, one requires a lens to form a virtual image. The lens of the human eye or the lens of a camera can be used. Thus, one can look at the light diffracted by the hologram and see the original object.

A second image is formed by light diffracted in a different direction from the virtual image. This is called a real image; it can be projected directly onto a screen and does not need a lens to form it. The real image must be projected onto some surface in order to view it.

There will also be some light transmitted directly through the hologram, without being diffracted. This undiffracted light will emerge from the hologram traveling in the same direction as the reconstructing beam. This undiffracted light is not of great interest. It simply represents a loss of some of the light that might have gone into one of the images. The two other distributions of light that correspond to the two images will be on opposite sides of the undiffracted beam. The relative orientations are as shown in Figure 18-2.

Before we describe how the holographic process operates, let us consider briefly the history of holography. Holography was first invented by Dennis Gabor in 1948. Gabor's motivation was to provide a new method of microscopy. The holographic process can involve magnification. The exact value of the magnification depends on the radii of curvature of the wavefronts used for recording the hologram and for reconstructing the images. Under reasonable conditions, the reconstructed image will be magnified by the ratio of the wavelengths of the light used for reconstructing the hologram and for producing it. Gabor's idea was to record holograms using x-rays and to reconstruct them using visible light. This would lead to a magnification by a factor of several thousand.

Application of holography to high-magnification microscopy was hampered by two factors: (1) the lack of a high-quality coherent x-ray source for recording the holograms, and (2) the lack of a bright coherent light source for reconstructing them. After some interest in the late 1940s, holography was relatively dormant until the advent of the laser. The laser provided a bright coherent source that can be employed both for recording and reconstructing holograms. It was soon recognized that holography using lasers could yield high-quality three-dimensional images, and beginning in the early 1960s, there was a great revival of interest in holography. This interest was spurred further by an advance developed by E. N. Leith and J. Upatnieks at the University of Michigan. In Gabor's original work, the reference beam and object beam used for recording the hologram were both incident in the same direction. Then, when the hologram was reconstructed, using a reconstructing beam traveling in the same direction as the original reference beam, the three beams were coincident. The real image, the virtual image, and the undiffracted transmitted beam were collinear in space. This made it difficult to distinguish the images.

The contribution of Leith and Upatnieks was the off-axis geometry, as shown in Figure 18-1. In this geometry, the reference beam and the object beam are incident on the recording medium at an angle to each other. When the hologram is reconstructed by light traveling in the same direction as the reference beam, the three resulting beams are separated in space. The undiffracted transmitted beam travels in the same direction as the reconstructing beam. The real image and the virtual image are produced in directions at an angle to the reconstructing beam, on opposite sides of it. The physical separation makes it much easier to view the images.

These developments made holography practical and useful. Many applications for holography have developed. Many additional methods and geometrical arrangements for producing holograms have been devised. The original application envisioned by Gabor for microscopy still has not been exploited, because there is not yet a good high quality source of x-rays suitable for recording holograms. Among other reasons, this lack has motivated efforts to develop x-ray lasers.

When the general arrangement of object and reference beams and the recording medium is as shown in Figure 18-1, the hologram is called a Fresnel hologram. A Fresnel hologram is formed when the recording medium is relatively near the object—more precisely, in the near field of the diffraction pattern of the object. In addition to Fresnel holograms, there are a variety of types of holography. Other vari-

ations in the recording process include the use of converging or diverging beams and also illumination of the recording medium with the object and reference beams incident from opposite sides. Holograms can be recorded that can be viewed in light reflected from the hologram or in light transmitted through it. It is beyond the scope of this chapter to consider in detail all the various types of holograms.

One geometry that is worth considering is the Fourier transform hologram. Such holograms generate plane wave amplitudes at the hologram, which is the Fourier transform of the object. An arrangement for forming a Fourier transform hologram is shown in Figure 18-3. There are other methods for producing Fourier transform holograms that do not require the use of a lens. The discussion here will show the essential features of the Fourier transform hologram. The Fourier transform hologram is restricted by the requirement that the object must lie within a single plane.

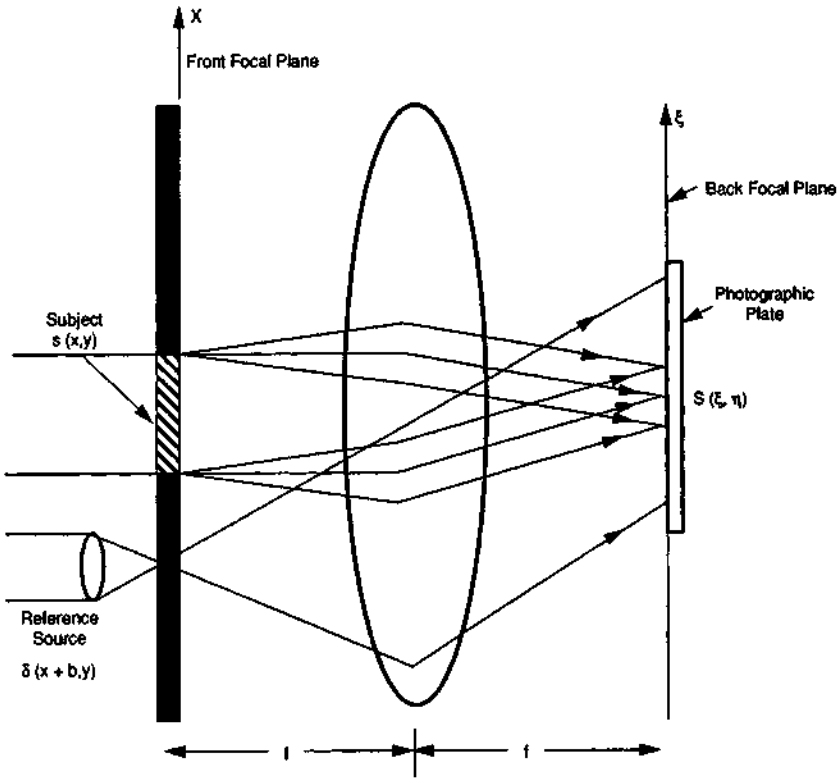


Figure 18-3 Arrangement for formation of a Fourier transform hologram. The focal length of the lens is f. The spatial coordinates in the front focal plane are x and y; those in the back focal plane are ξ and η. The light distribution of the subject is denoted by the function s; that in the transform plane is S. The reference beam is incident through a pinhole and can be represented by a δ-function.

Thus, the objects for Fourier transform holograms tend to be photographic transparencies. The method is less applicable to subjects that have an extended depth. Fourier transform holograms are of interest for use in optical pattern recognition and for optical computer memories. For such applications, the presentation of the data in a single plane is not a disadvantage.

The Fourier transform hologram utilizes the phenomenon that a lens converts a distribution of light in one of its focal planes into the Fourier transform of that distribution in the other focal plane. Thus, if the distribution of light in the first focal plane of the lens is given by the function $s(x, y)$, the distribution in the second focal plane will be the Fourier transform of $s(x, y)$. (The two-dimensional Fourier transform, $S(\xi, \eta)$, of the function $s(x, y)$ is defined by the equation

$$S(\xi, \eta) = \iint s(x, y) \exp i(\xi x + \eta y)\, dx\, dy \tag{18.1}$$

where x and y are the coordinates in the first focal plane of the lens, ξ and η are the coordinates in the second focal plane, and the integration is over all space.) In the reconstructing process, another Fourier transform is performed by a second lens. The net effect of the two Fourier transforms restores the original distribution of light, $s(x, y)$, in the back focal plane of the second lens, except that the image is upside down and backward. The same process converts the conjugate image, which is already reversed, into the same light distribution as the original wavefront. Thus, the Fourier transform hologram yields two images, reversed in direction from each other, on opposite sides of the undiffracted beam. An advantage of a Fourier transform hologram is that the angular spread of the reconstructing beam may be large [1]. The Fourier transform hologram has applications in the processing and manipulation of data presented in the form of distributions of light intensity.

B. The Holographic Process

We shall now consider how the holographic process operates. We begin by giving a simple heuristic picture of how a hologram is formed and how the reconstruction produces an image of the original object. Consider an object that is a point source, a pinhole. When a coherent plane wave is diffracted by a pinhole, one obtains a spherical wave. When the spherical wave is combined with a plane reference wave, the intensity distribution will consist of annular zones of constructive and destructive interference. This is shown in Figure 18-4. This distribution resembles a Fresnel zone plate. As one knows from elementary optics, the Fresnel zone plate consists of a series of concentric circles, with radii proportional to the square roots of the even integers. The zone plate acts as a diffraction grating with focusing properties. It operates like a lens with both positive and negative focusing characteristics. If collimated light falls on the zone plate, some of the light is diffracted so as to converge to a point. This is a real image and represents the positive focusing capability of the zone plate. Some of the diffracted light will appear to diverge from a point source behind the zone plate. This represents the negative focusing properties of the zone plate.

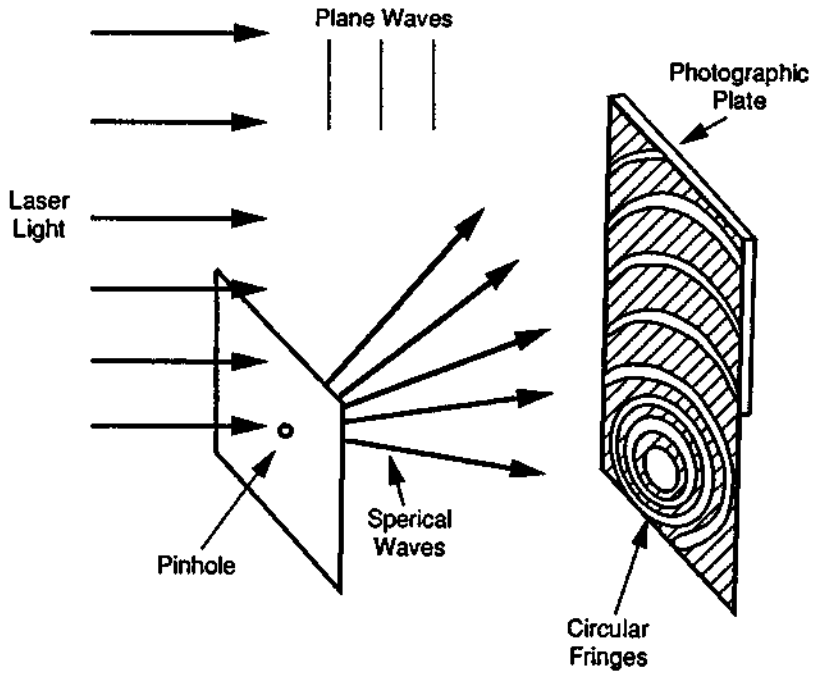


Figure 18-4 Formation of a Fresnel zone plate. (From W. E. Kock, L. Rosen, and J. Rendeiro, *Proc. IEEE* **54**, 1599 (1966).)

In accordance with the positive focusing properties, the zone plate of concentric circles formed as shown in Figure 18-4 will act as a lens and will refocus a plane wave to a point, which corresponds to an image of the pinhole, admittedly a very simple object. This leads to an understanding of the reconstruction process. The Fresnel zone plate diffracts two beams of light. One is a spherical wave that appears to diverge from the position of the pinhole. This is a virtual image of the pinhole. An observer looking at the light will see a bright point of light at the position of the pinhole. The second beam is a converging bundle of light rays, which forms a real image of the pinhole. If a piece of paper is placed at the focus, a bright spot of light will be observed on the paper. This is shown in Figure 18-5.

A more complicated object can be considered as an ensemble of point objects. The light from each point produces its own zone plate by interfering with the reference beam. All the zone plates add together to form a complicated interference pattern. In reconstruction, the same reference beam is used to re-create all the points comprising the object. Thus, when one views the reconstruction, one sees the same image, including the three-dimensional nature, as one would have seen by looking at the original object directly.

The foregoing discussion explains intuitively how the holographic process works. It is not a proof. We now proceed with a more accurate description, which requires use of more mathematics.

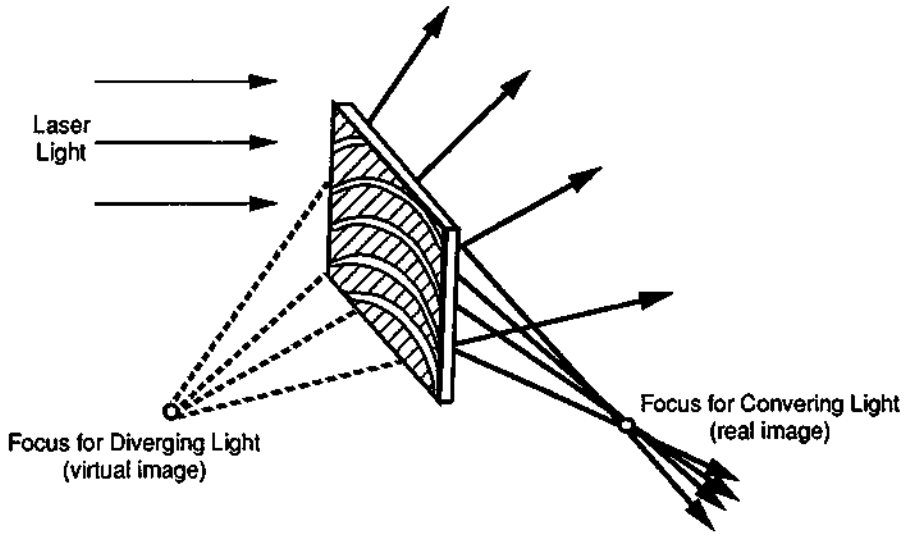


Figure 18-5 Reconstruction of a point image from a Fresnel zone plate. (From W. E. Kock, L. Rosen, and J. Rendeiro, *Proc. IEEE* **54**, 1599 (1966).)

We refer to Figure 18-6 for this description. Figure 18-6 shows the recording of a hologram. It contains the essential features as shown in Figure 18-1, namely, a plane reference beam and an object beam intersecting at an angle θ. The exact details of where the beams come from (laser, beam splitter, etc.) have not been included. Also, the object is shown as a two-dimensional transparency. At the plane of the photographic film, the light will have a distribution $S(x, y)$ that is characteristic of the object. We represent this as

$$S(x, y) = S_0(x, y) \exp i\left[\omega t + \frac{2\pi z}{\lambda} + \phi(x, y)\right] \tag{18.2}$$

where x and y are the coordinates in the plane of the film, z is the direction of propagation of the light, ω is the angular frequency of the light, S_0 is an amplitude function, representing the instantaneous distribution of the light amplitude in the plane of the film, and ϕ is a function describing the phase of the light. We note that S_0 and ϕ are functions of the spatial coordinates, and that S is a complex number.

In comparison, a wave representing an electric field, oscillating at angular frequency ω, can be denoted as

$$E(x, y, t) = E_0(x, y) \cos[\omega t + \phi(x, y)] \tag{18.3}$$

where t is the time, E_0 (the amplitude) is the maximum value of the electric field, and ϕ (the phase) represents the status of a particular point as it moves through its cyclic variation.

In this discussion of holography, we shall employ complex notation, using the equality $e^{ia} = \cos a + i \sin a$. The amplitude function just defined can be regarded as

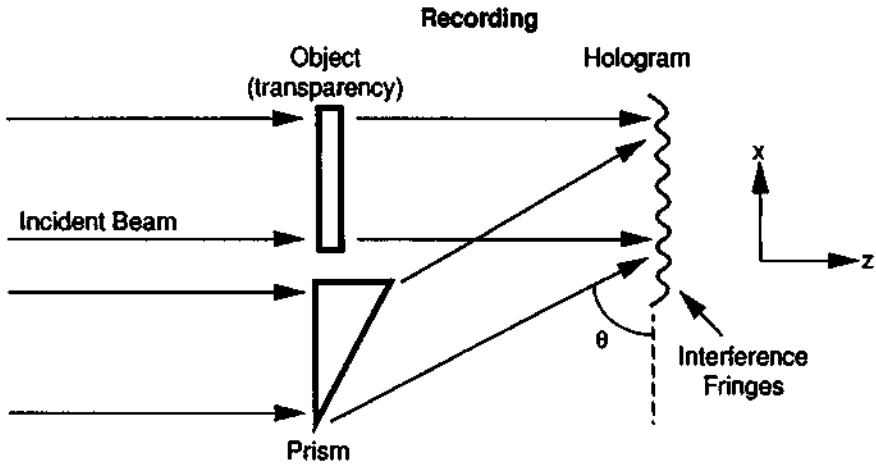


Figure 18-6 Recording of a hologram.

the amplitude of the wavelike motion of the light. Irradiance (the power per unit area delivered by the wave) is proportional to the square of the absolute value of the amplitude. All optical detectors, including photographic film, respond to irradiance, not to amplitude directly. The diffraction phenomena that give rise to holography must be described in terms of the wave nature of the light and, hence, in terms of the amplitude and phase functions. When we describe the recording process in the film, we must square the amplitude.

The instantaneous amplitude of the reference beam $R(x, y)$, a plane wave, is

$$R(x, y) = R_0 \exp i\left(\omega t + \frac{2\pi x}{\lambda} \cos \theta + \frac{2\pi z}{\lambda} \sin \theta\right) \tag{18.4}$$

The amplitude function R_0 is a constant because the reference beam is a plane wave, with constant amplitude across the wavefront. It travels at an angle θ to the x axis. This is shown by the constant phase factor, $(2\pi \cos \theta)/\lambda$. We will define $\alpha = (2\pi \cos \theta)/\lambda$.

In the representation of both the object wave and the reference wave, there is a factor $e^{i\omega t}$. Because the recording process involves an average over many cycles of the optical field, this term will yield only a constant multiplying factor, which we will neglect from now on. In addition, if the plane of the film is relatively thin, the value of z will not change much throughout the film. Without loss of generality, we may take $z = 0$. We shall later consider what occurs if the emulsion of the film cannot be regarded as thin.

With these approximations, we have, for the representations of the object and reference waves, respectively,

$$S(x, y) = S_0(x, y)e^{i\phi(x, y)} \tag{18.5}$$

$$R(x, y) = R_0 e^{i\alpha x} \tag{18.6}$$

Because the irradiance of the light wave is proportional to the square of the absolute value of the amplitude, if the object beam alone were incident on the film, the irradiance would be

$$|S(x, y)|^2 = [S_0(x, y)]^2 \tag{18.7}$$

neglecting a constant proportionality factor.

This pattern of irradiance would be recorded in a conventional photograph. The photographic film records the irradiance, averaged over time. Thus, all the factors involving $e^{i\omega t}$ will drop out. The pattern on the photographic film will preserve the amplitude function S_0, but all information about the phase function ϕ will be lost.

Now consider the result when the film is exposed to the irradiance distribution resulting from the sum of the object and reference beams. This distribution, $I(x, y)$, recorded by the film will be (dropping for the moment the functional notation for R, S, R_0, S_0, and ϕ)

$$\begin{aligned} I(x, y) &= |R + S|^2 = |R|^2 + |S|^2 + RS^* + R^*S \\ &= R_0^2 + S_0^2 + R_0 S_0 (e^{i\phi} e^{-i\alpha x} + e^{-i\phi} e^{i\alpha x}) \\ &= R_0^2 + S_0^2 + R_0 S_0 [\cos(\alpha x - \phi) + i \sin(\phi - \alpha x) \\ &\quad + \cos(\phi - \alpha x) - i \sin(\phi - \alpha x)] \\ &= R_0^2 + S_0^2 + 2R_0 S_0 \cos(\alpha x - \phi) \end{aligned} \tag{18.8}$$

where the notations S^* and R^* denote the complex conjugates of S and R, respectively. According to Equation (18.8), the film records a pattern that has the sum of the squares of the amplitude terms, plus an oscillating cosine term. This describes a pattern of fringes. The terms R_0 and S_0 yield a relatively slowly varying background that is modulated by the fringes, which have high spatial frequency. We note that the recording contains information about both the amplitude function $S_0(x, y)$ and the phase function $\phi(x, y)$.

The next step in the process is the development of the photographic film. We digress here to review the characteristics of the response of photographic film, which are generally described by the so-called H–D curve. This curve is represented in Figure 18-7. It is a plot of the optical density of the developed film as a function of the logarithm of the exposure (in units of energy per unit area or, equivalently, in irradiance per unit area times the exposure time). The optical density D is defined by

$$D = -\log_{10} T \tag{18.9}$$

where T is the transmission of the developed film. Thus, optical density equal to unity means that the transmission is 0.1. Generally, the response of the film depends only on the total energy per unit area incident on the film. This phenomenon is called reciprocity. It does not matter whether the energy is delivered slowly at low power or quickly at high power. All that matters is the total energy density delivered. Reciprocity does break down for very short exposures, but it holds over a wide range

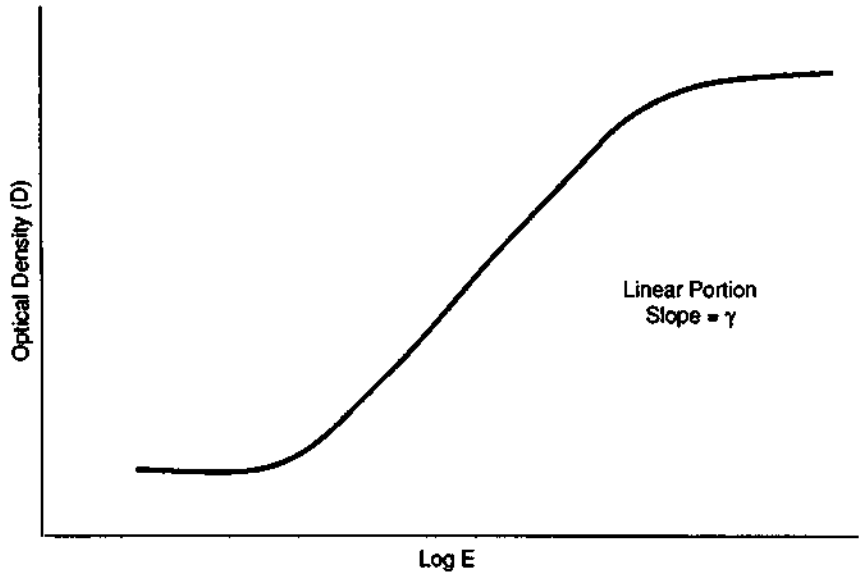


Figure 18-7 Response of photographic film: Optical density D as a function of exposure E, which is in units of energy per unit area.

of exposure times, which cover the range of interest for most photographic and holographic purposes.

According to Figure 18-7, when the recording medium is a conventional photographic film, the result of exposure and development is a change in transmission of the film, with the increase in optical density being greatest in regions where the irradiance was highest. Thus, the fringes in the light wave are converted into fringes in optical density, and the interference pattern is frozen as an absorbing pattern of light and dark areas in the film.

The form of the curve in Figure 18-7 is of interest. At low exposures, there is relatively little response. This represents the background fog level of unexposed film. At higher exposures, the optical density increases with increasing exposure. Near the right side of the figure, the curve stops increasing and approaches a constant value. In this region, the response of the film saturates and the film is overexposed. Generally, for holographic purposes, one desires to work in the region where the H–D curve rises linearly. The slope of the curve in the linear region is conventionally termed γ. The value of γ is a function of the particular type of film and of the development parameters, including the type of developer and the time and temperature used in the development. The values of γ for specific conditions are often supplied by the manufacturers of photographic film. For many types of film, γ may be approximately equal to 2. In the linear portion of the H–D curve, the transmission will be approximately proportional to the exposure to the minus γ power. Thus, we have the relation that the transmission T is proportional to $E^{-\gamma}$, where E is the

exposure (energy per unit area). In a negative, γ is greater than zero, and the transmission decreases as exposure increases, as we expect. If the film is printed as a positive, we may consider γ to be a negative quantity.

The transmission just described is that for irradiance falling on the film. To describe the holographic process, we need the transmission T_a for the light amplitude, given by

$$T_a = T^{0.5} \tag{18.10}$$

Thus, T_a is proportional to $E^{-\gamma/2}$. With proper control of the development parameters, such as time and temperature, so as to give $\gamma = -2$ (i.e., printing as a photographic positive), one may have a situation where T_a is proportional to E. Thus,

$$T_a = AE \tag{18.11}$$

where A is a constant of proportionality, the exact value of which depends on the conditions during development. If the development yields a value of γ different from -2, additional terms would have to be added to the equations presented in what follows. This would yield additional images that could be viewed. Such additional images in fact are often seen in holograms.

This discussion implies development as a photographic positive. If a photographic negative is printed, the regions of high and low transmission would be interchanged. The light and dark fringes would be reversed, but essentially the same interference pattern would still be present in the film. The holographic process would still operate as will be described.

The result of the properly controlled exposure and development procedures is a piece of developed film with amplitude transmission given by

$$T_a = A[R_0^2 + S_0^2 + 2R_0S_0\cos(\alpha x - \phi)] \tag{18.12}$$

where we have used the result of Equation (18.8) in (18.11). If we now reilluminate this film with the reference beam alone, the light pattern transmitted by the film will be the reference beam multiplied by the film transmission, namely,

$$R(x, y)T_a(x, y) = AR_0e^{i\alpha x}[R_0^2 + S_0^2 + 2R_0S_0\cos(\alpha x - \phi)] \tag{18.13}$$

where we are still suppressing the functional notation for R_0, S_0, and ϕ. Then, the distribution of light transmitted by the film is

$$\begin{aligned} RT_a &= AR|S|^2 + AR|R|^2 + A|R|^2S + AR^2S^* \\ &= ARS_0^2 + ARR_0^2 + AR_0^2\mathrm{S} + AR^2S^* \end{aligned} \tag{18.14}$$

where we have used the first part of Equation (18.8). Of the terms in this equation, the first two taken together, $AR(R_0^2 + S_0^2)$, are equivalent to the reconstructing beam multiplied by the relatively slowly varying factor $A(R_0^2 + S_0^2)$. This, therefore, represents undiffracted light traveling in the same direction as the reconstructing beam. Figure 18-8 shows this beam as the beam transmitted straight through the hologram.

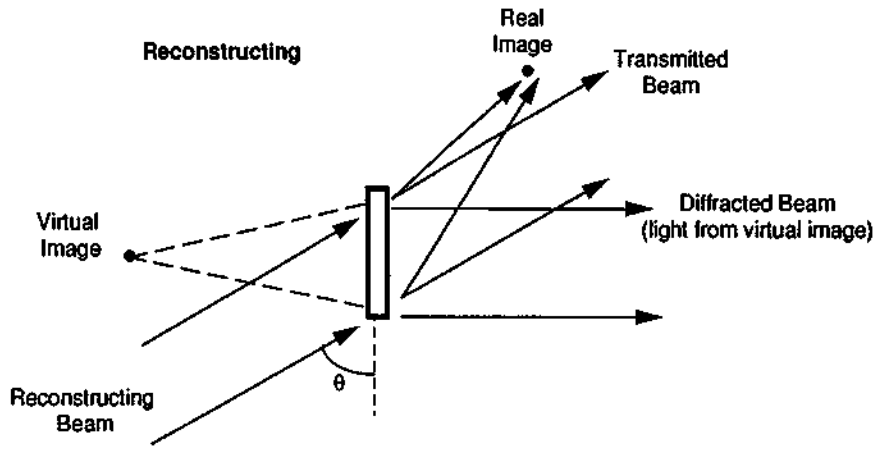


Figure 18-8 Reconstruction of a hologram.

The next term

$$AR_0^2S = AR_0^2S_0(x, y)e^{i\phi(x, y)} \tag{18.15}$$

can be recognized as the light distribution from the original object, multiplied by the constant term AR_0^2. The functional notation has been restored here. The constant multiplier does not change any of the essential features of the image; it will only change the brightness of the image. Thus, this term represents a light distribution that is identical to what would have come from the original object. Viewing this light gives a view of the original object. Both the amplitude and phase functions are present, so that all the information about the object has been preserved. This includes the three-dimensional nature of the object. This is a virtual image. The angular separation between the light in the virtual image and the undiffracted light is shown in Figure 18-8.

The light from the virtual image is diverging, apparently from the position where the original object was located during recording. Thus, one must use a lens to collect and refocus this light in order to see the image. The image will appear to be hanging in space behind the hologram. To view the virtual image, one should relax the eye and look through and behind the hologram.

Figure 18-9 shows two photographs of the reconstruction of a virtual image from a hologram. The photographs (two dimensional, of course) were taken at two different angles of viewing the reconstructed image. The three-dimensional nature of the image is apparent in the parallax that may be seen between the two different views.

The emergence of this term, $AR_0^2S_0(x, y) \exp i\phi(x, y)$, which represents the light from the object, is not merely an artifact of the mathematics. The term represents real light that emerges from the hologram during reconstruction, and which may be seen by the human eye. The presence of this light, with a distribution identical to

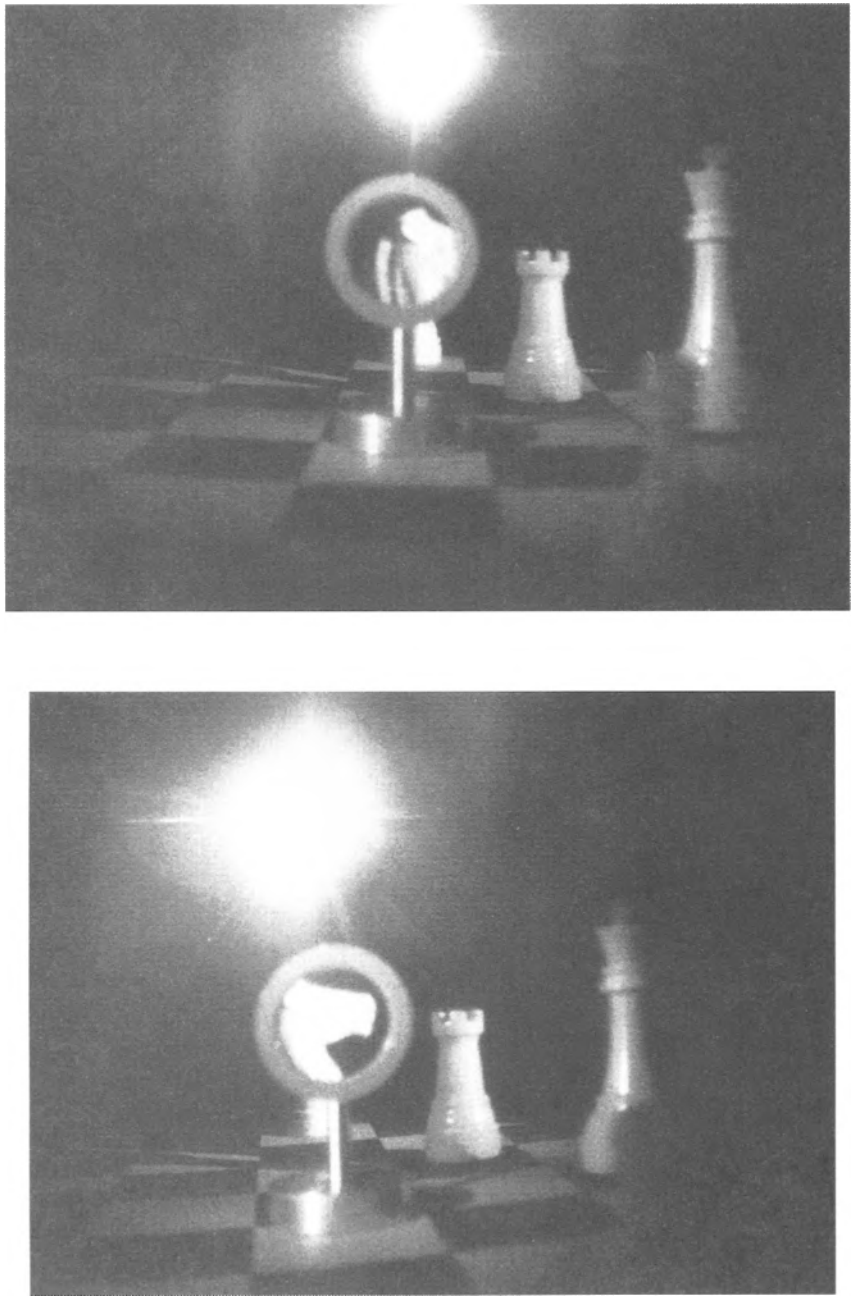


Figure 18-9 Photographs of the reconstruction of a hologram. The two photographs, taken at slightly different angles, show the parallax arising from the three-dimensional nature of the hologram. The bright spot near the top is the undiffracted beam.

light that came from the original object, is the essential point of the holographic process. Information about the object is stored in the fringe pattern stored in the film. When the reconstructing beam is diffracted by the fringes, the emerging light is distributed in such a way that the image of the original object is re-formed.

The final term in Equation (18.14),

$$AR^2S^* = AR_0^2S_0(x, y) \exp i[2\alpha x - \phi(x, y)] \tag{18.16}$$

represents light that has amplitude and phase functions preserved. However, the negative sign in front of the phase factor means that this light is converging, not diverging. The light can form an image on a flat surface placed in the plane at which the light converges. Thus, this represents the real image. The angle θ' at which this light is traveling with respect to the x axis is given by the angular factor $e^{i2\alpha x}$. We have

$$\cos \theta' = \frac{2\alpha\lambda}{2\pi} = 2 \cos \theta \tag{18.17}$$

Thus, $\cos \theta' > \cos \theta$, where θ is the angle between the x axis and the reconstructing beam. Hence, the real image appears on the opposite side of the undiffracted beam (the opposite side from the virtual image). This is also shown in Figure 18-8.

In the holographic recording process, every point on the object sends light to every point on the hologram. Thus, the fringe pattern is spread over the entire area of the hologram, so that information about the entire object is stored over the entire hologram. As a result, the entire scene may be reconstructed from any small area of the hologram (or at least from any area large enough to store a reasonable number of fringes). Holograms have the property (in contrast to conventional photography) that if a hologram is cut in half, one has two holograms, each containing the entire picture.

C. Hologram Types and Efficiency

So far we have discussed the use of a thin recording medium for holograms. Such a medium will record the contours of light intensity in the plane of the medium. But if the photosensitive medium is relatively thick, the contour surfaces of light intensity will be recorded throughout the thickness of the medium. Figure 18-10 indicates the distinction between thick and thin recording media. On the thick hologram, there will be a number of fringes throughout the thickness of the material, as one passes through it in a direction perpendicular to the surface. In a thin hologram, one would encounter at most a single fringe in passing through the material. The distinction between thick and thin is not an absolute distinction based on the physical thickness of the material. A piece of film of a given thickness could be either a thick or thin hologram depending on the method by which the hologram was recorded. One can show that if the z-dependent terms in Equations (18.2) and (18.4) were retained, the treatment would be modified so as to yield the following result in place of Equation (18.8):

$$|R + S|^2 = R_0^2 + S_0^2 + 2R_0S_0F \tag{18.18}$$

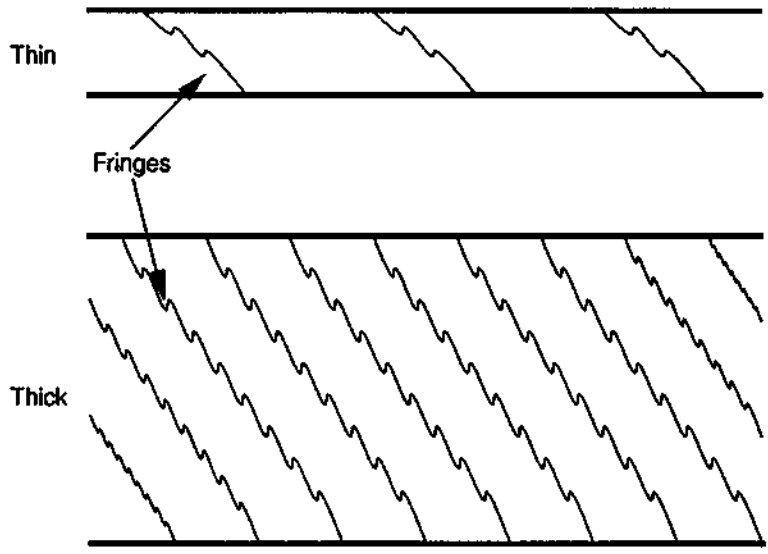


Figure 18-10 Illustration of thin and thick holographic recording media.

where the additional factor F is given by

$$F = \cos(\alpha x - \phi)\cos(\beta z - kz) - \sin(\alpha x - \phi)\sin(\beta z - kz) \qquad (18.19)$$

with $k = 2\pi/\lambda$ and $\beta = (2\pi \sin\theta)/\lambda$. Thus, in the thick hologram, the positions of the contours of intensity will depend on both the x and z position throughout the film, whereas in the thin hologram, they depend only on the x position. Thick holograms are sometimes called volume holograms because the fringe pattern is stored throughout the entire volume of the recording medium.

The properties of thin and thick holograms are different. Thick holograms are particularly useful for increasing the efficiency of the reconstruction process. The efficiency is defined as the ratio of the amount of light in the image to the amount of light in the reconstructing beam. High efficiency contributes to brighter, more easily visible holographic images. The values of efficiency that are available with different types of holograms will be discussed later.

So far, we have considered only the case in which the transmission of the recording medium is varied by the incident light to record the fringe pattern. Such holograms operate by absorption of some of the reconstructing light. These holograms are called absorption holograms or, equivalently, amplitude holograms. The most common amplitude hologram is based on the use of photographic film, which originally contains silver halide grains, which are converted to free silver metal as a result of the exposure and development processes. The free silver is most dense where the light intensity was highest. The free silver is absorbing, so that the fringe pattern is stored in the film as a variable pattern of absorption.

There are other methods whereby the recording can be made through changes in the index of refraction of the recording medium. Such holograms are called phase

holograms. The recording of the fringes produces regions in which the optical path length through the medium is changed, because of the change in the index of refraction. With such holograms, there is no absorption of the reconstructing light, so that phase holograms can have higher efficiency than amplitude holograms.

There are several types of media suitable for recording phase holograms. These include photographic film, in which an amplitude hologram is originally recorded; the film is then bleached through chemical processes. In the bleaching, the silver grains that cause absorption are removed, and the medium appears transparent. But the index of refraction has been changed in regions where the silver grains were developed, that is, in regions where the light intensity was high. Thus, the interference pattern is still stored in the hologram as a variable pattern of the index of refraction.

There are other media that produce phase holograms directly without the necessity of a bleaching step. One such medium is dichromated gelatin, in which a gelatin emulsion is impregnated with a chemical like ammonium dichromate. Interaction of the light with the ammonium dichromate tends to tan the gelatin and changes the index of refraction. Thus, dichromated gelatin films are directly used to produce phase holograms without having to employ an intermediate bleaching step. Dichromated gelatin has many desirable features, including low absorption, high resolution, and the potential for high brightness. However, dichromated gelatin films have several disadvantages. They are less sensitive to light than conventional photographic emulsion. Their spectral sensitivity does not extend into the orange or red, so that helium–neon lasers cannot be used to make holograms on dichromated gelatin films. Dichromated gelatin is not commercially available and must be prepared by the user. We shall return to this point later.

There is another distinction in types of holograms, the distinction between transmission and reflection holograms. So far, we have discussed only transmission holograms. The fringes in light intensity are formed perpendicular to the plane of the film. The object and reference beams are incident from the same side of the film, and the image is viewed by transmission of the reconstructing beam through the developed film. But for a reflection hologram, the object beam and reference beam are incident from opposite sides of the original film. The interference fringes may be nearly parallel to the plane of the film. A number of fringes will be formed through the thickness of the material. The recorded interference pattern resembles a stack of closely spaced planes, roughly parallel to the surface of the film, as shown in Figure 18-11. This stack acts a reflective filter. This can be either an amplitude filter (for example, reflecting planes of metallic silver grains) or a phase filter, that is, planes in which the refractive index varies. Of necessity, a reflection hologram must be a thick hologram.

During reconstruction, the illuminating light is incident from the same side of the film as the observer, and the image is viewed in reflection from the film. This is also shown in Figure 18-11. Polychromatic light or white light may be used for reconstruction of such holograms. Thus, these holograms are often called white light holograms. Different spectral components will be reflected at different angles, so that the apparent color changes as the viewing angle changes. These holograms offer the convenience that viewing the image does not require a laser. Thus, the

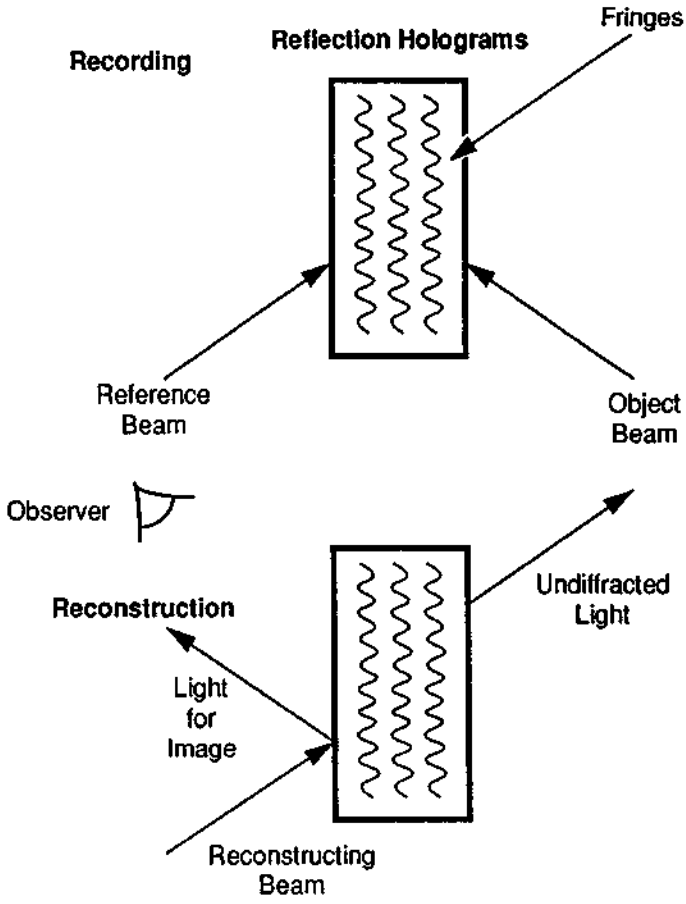


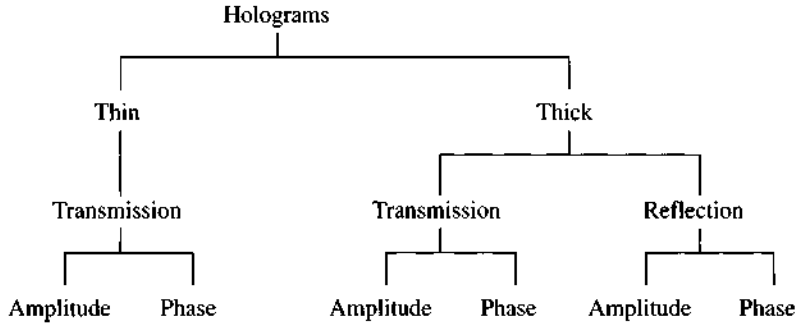
Figure 18-11 Recording and reconstruction for reflection holograms.

holograms that one commonly sees for artistic or display purposes or on credit cards are reflection holograms. White light reflection holograms are best viewed with a single point source of light, such as the sun or a slide projector.

As this discussion shows, the use of a laser is not always necessary to view a hologram. And as we shall describe later, holograms may be replicated without use of lasers. But lasers are necessary to record high-quality holograms, at least the original master hologram. Laser technology was required before holography could progress and become practical.

We now review the relation of these different types of holograms. Table 18-1 summarizes the situation. The first distinction is between thin and thick holograms, depending on the relation of the fringe spacing to the thickness of the recording medium. Thick (or volume) holograms can be either reflection or transmission holograms. Thin holograms are necessarily transmission holograms. Finally, each of these three types may be recorded as either a phase or absorption hologram.

Table 18-1 **Classification of Holograms**



Now let us consider the efficiency, defined as the amount of light diffracted into the image divided by the amount in the reconstructing beam. A full discussion of the efficiency of all types of holograms is beyond the scope of this chapter. However, the maximum theoretical values of efficiency for several important types are tabulated in Table 18-2. This table gives the maximum possible efficiency for each of the different types of holograms defined in Table 18-1. We see that the maximum efficiency for absorption holograms is relatively low, around 7 percent at most. But the efficiency of phase holograms is relatively high, and it always exceeds that of the corresponding type of amplitude hologram. The reason is obvious: amplitude holograms rely on absorption of light and thus are wasteful of the energy that is used to illuminate them for reconstruction, whereas phase holograms are nearly transparent and use the reconstructing light more effectively. The theoretical efficiency defined in Table 18-2 has been approached in practice for each of the types of hologram we have defined. In particular, for thick transmission phase holograms in dichromated gelatin, efficiency in excess of 90 percent has been obtained experimentally.

In general, one desires high efficiency. The reconstructed image is brighter and the requirements for laser power in the reconstructing beam are reduced when efficiency is high. For a thin amplitude hologram, at most about 6 percent of the reconstruction light can appear in the image, whereas for thick phase holograms, most of the reconstructing light can contribute to the image. Thus, although they are more difficult to produce, thick phase holograms are often used for applications.

Table 18-2 **Maximum Efficiency of Various Types of Holograms**[a]

Absorption holograms			Phase holograms		
Thin transmission	**Thick transmission**	**Thick reflection**	**Thin transmission**	**Thick transmission**	**Thick reflection**
6.25%	3.7%	7.2%	33.9%	100%	100%

[a]Maximum theoretical values.

D. Practical Aspects of Holography

Now we will discuss some of the practical aspects of holography. The light from the laser must be expanded and collimated to cover a sufficiently large area, so as to include the entire scene of which the hologram is to be made. Because the beams from many common lasers used in holography have diameters around 1.5 mm, one often uses a beam expander, as was shown in Figure 18-1. The light must also be spatially coherent over the entire scene. Otherwise, the contrast in the fringes will be degraded. Hence, one usually employs a laser operating in the TEM_{00} mode to ensure good spatial coherence.

To make holograms successfully, one should use a laser with good monochromaticity. This is equivalent to saying that the temporal coherence should be high. The coherence length of the laser must be greater than the depth of the scene being rendered. The coherence length is inversely proportional to the width of the frequency spectrum of the laser. The frequency spectrum is broadened by the presence of more than one longitudinal mode in the laser. If a helium–neon laser operates in a single longitudinal mode, its spectral width at half-intensity is of the order of 10^6 Hz. The coherence length is then of the order of 10^4 cm, which would be much greater that the depth of the scene in practical holographic applications. One could make good holograms for objects with depths much less than 10^4 cm.

But if more than one longitudinal mode is present in the laser output, the frequency spread is increased and the coherence length is decreased. If two different longitudinal modes with slightly different wavelengths traverse a holographic setup with a finite depth of field for the object, the two modes will be somewhat out of phase when they reach the recording medium. This will reduce the visibility of the fringe pattern and decrease the quality of the hologram. To obtain high-quality holograms, one must have

$$D \ll \Delta l = \frac{c}{\Delta f} \tag{18.20}$$

where Δl is the coherence length, Δf is the frequency spread of the laser, and D is the maximum difference in path length for light reaching the recording medium. The maximum value of D may be interpreted as the maximum depth of field that can be preserved in the hologram or the maximum path difference that may be allowed between the object and reference beams. If the laser is L cm long, and if N longitudinal modes are present in its output, one has

$$\Delta f = \frac{(N-1)c}{2L} \tag{18.21}$$

Hence, one must have

$$D \ll \frac{2L}{\mathrm{N}-1} \tag{18.22}$$

If a multimode laser is employed, the depth of field for the object will be small, and the path lengths that the object and reference beams travel must be closely matched. To produce a hologram of an object with a depth greater than a few centimeters, one must employ a laser operating in a single longitudinal mode. Otherwise, the interference pattern will be washed out because of dephasing between the different frequency components at opposite ends of the laser spectrum. If a single-mode laser is used, the restrictions on path length are much reduced. One may simply use the width Δf of a single longitudinal mode in Equation (18.20).

If there are N separate longitudinal modes in the laser output, the fringe visibility V is given by

$$V = \left| \frac{\sin(N\pi D/2L)}{N \sin(\pi D/2L)} \right| \tag{18.23}$$

This parameter is plotted in Figure 18-12 as a function of the path difference D. When several longitudinal modes are present, the fringe visibility decreases if the path difference becomes greater than a few centimeters. This has the effect of reducing the possible depth of field for the object. If a laser with several longitudinal modes in its output is to be used, any path differences in the apparatus must be reduced to less than a few centimeters in order to preserve good fringe quality. With a single-mode laser, the depth of field can be much greater. The disadvantage of single-mode lasers is that their output power is lower than that of multimode lasers.

Another very important factor is the elimination of any relative motion between the components in the holographic apparatus, during the time that the hologram is

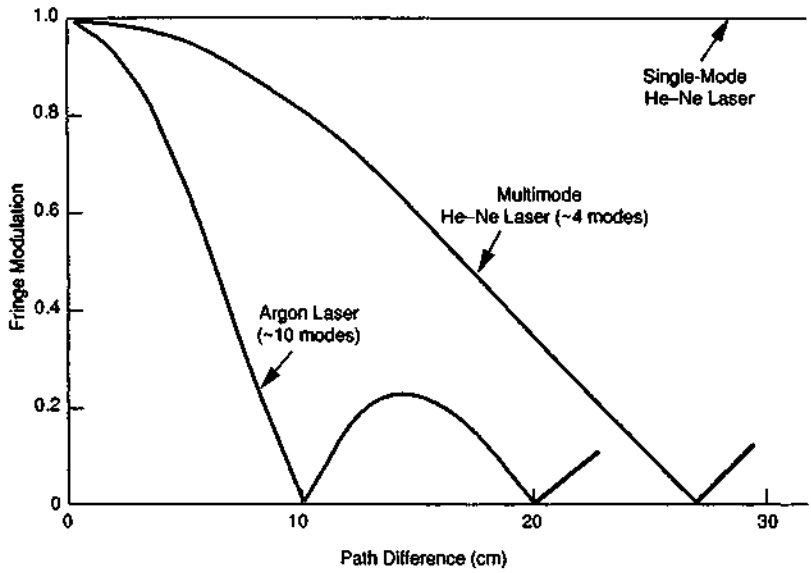


Figure 18-12 Fringe visibility in a hologram as a function of path difference in the holographic apparatus for lasers with different numbers of longitudinal modes.

recorded. If the distance between any of the components changes by one-half the wavelength of the light, a fringe will change from dark to bright. Thus, any positional change that alters the optical path will degrade the fringe pattern. A change of half a wavelength during recording will do substantial harm to the quality of the hologram. One generally tries to hold any positional changes to less than one-tenth wavelength of the recording light. Thus, one requires rigid mounts and platforms, stable positioning, and elimination of vibration. Effective holographic platforms have been constructed using commercially available metal honeycomb tables or using granite slabs, as have been described in Chapter 5. Such tables are mounted on vibration isolation platforms to decouple them from the vibration environment. The tables offer high rigidity, that is, they tend to move as a whole, and not to deform so as to introduce relative motion between components mounted on them. If the platform simply moves as a whole without deformation, it will not affect the relative positions of the components and, thus, will not affect the interference pattern.

The problems with vibration isolation are worsened by another factor. The fine-grained films required to record the closely spaced interference fringes tend to be slower than the high-speed films that are now commonly encountered in conventional photography, so that holographic exposures tend to take longer than the time to make a conventional photograph. Thus, the quality of the vibration isolation platform is important. Commercially available packages are generally adequate.

Despite what has been said about elimination of motion, holograms of moving objects can be made by using very short pulses of light from Q-switched lasers. The relative position of the components does not change much during the pulse of laser light, which typically has a duration in the regime of tens of nanoseconds. This is true even if the object is moving rapidly. As an example, holograms have been made of bullets in flight, with a Q-switched ruby laser with 10 nsec pulse duration as a source. An example will be presented in the next chapter.

Air currents and temperature changes can vary the optical path length between the components by varying the index of refraction of the air. This can degrade the quality of the holograms just as much as physical motion would. Often, holographic work is done in a temperature-stabilized environment with shields to eliminate air currents.

The most common recording medium for holographic recording has been photographic film. Table 18-3 lists properties of some of the emulsions used in commonly available films suitable for holographic purposes. The table is representative of what is available and not a complete list, because new types of films are continually being developed. For each emulsion, the table gives its relative sensitivity in terms of the energy per unit area required to yield an optical density of unity, for several important laser wavelengths in the visible portion of the spectrum. We note that it is desirable for this parameter to be as low as possible, in order to make the exposure time short so as to minimize the effects of vibration. The table also presents data on the resolution of the emulsions, expressed in the number of lines per millimeter that can be recorded. This value can be high, in excess of several thousand, for the fine-grained emulsions listed here. This is necessary to record the closely spaced inter-

Table 18-3 **Photographic Emulsions for Holography**

Emulsion	Sensitivity (erg/cm^2) at wavelength (nm)				Resolution (lines/mm)
	514	532	633	694	
649F	800	1000	900	>5000	>2000
120/SO-173	—	—	400	400	>2000
125	50	100	—	—	1250
131/S0-253	25–35	20–30	5–8	>1000	1250
SO-243	—	—	5–10	—	500
8E56	350	300	—	—	5000
8E75	—	450	150	100	5000
10E56	20	20	—	—	3000
10E75	120	60	20	20	3000
14C75	—	—	5–10	—	1500
649F (bleached)	800	1000	900	>5000	>2000

ference fringes. Although other photographic emulsions with better sensitivity are available, they tend to be coarse-grained and do not offer adequate resolution to record the fringe patterns. For the emulsions listed in Table 18-3, there is still some trade-off between sensitivity and resolution, with the highest-resolution emulsions tending to require higher exposures. In the table, the first ten entries represent use as amplitude media, so that the maximum efficiency would be in the 4–6 percent range. The last entry, bleached 649F film, represents use as a phase medium, so that the efficiency could be in the 33 percent range. These films are available either with or without antihalation backings. The backings help reduce the presence of ghost images in the hologram, and backed films are often used for recording transmission holograms. For reflection holograms, the object and reference beams must be incident from opposite sides of the film, and so unbacked film must be used.

When one uses photographic film for recording holograms, one usually provides a total exposure that lies within the linear response of the film, in order to obtain maximum variation of optical density in the recorded fringes. This means that the relatively constant reference beam should provide an exposure that lies near the center of the linear portion of the H–D curve (Figure 18-7). This in turn means that the reference beam, at the position of the film, should by approximately five to ten times as intense as the object beam. This ratio is often used to give maximum fringe modulation.

The developing process for photographic film used in amplitude holograms is conventional, using standard procedures and familiar chemicals, developers, fixers, and so forth. The processing for photographic film used as phase holograms is nonconventional. One procedure for producing a bleached phase hologram is given in Table 18-4. This process, developed by the Eastman Kodak Company, yields high-quality phase holograms from photographic film. After exposure, one processes the film according to the steps in the table, instead of using conventional development.

Table 18-4 **Bleaching Process for Phase Holograms**

Step	Material[a]	Time	Temperature (°F)
Develop	Solution A plus Solution B (mix equal parts just before use)	5–8 min	75
Stop	Solution C	15 sec	70–80
Rinse	Running water	1 min	70–80
Bleach	Solution D	2 min	70–80
Wash	Running water	5 min	70–80
Dry	—	As needed	70–80

[a]Solution A: 750 ml water, 8 gm sodium sulfite, 40 gm pyrocatechol, 10 gm sodium sulfate, plus water to make 1 liter.
Solution B: 750 ml water, 20 gm sodium hydroxide, 100 gm sodium sulfate, plus water to make 1 liter.
Solution C: 1 liter water plus 48 ml acetic acid (28%).
Solution D: 1 liter water, 9.5 gm potassium dichromate, 12 ml concentrated sulfuric acid.

Because photographic film is easily available and its technology is well developed, it is the medium most often used for holography. Sometimes other considerations dictate the use of other materials. Table 18-5 lists some other recording media that have been employed for holographic applications. The table shows the wide range of materials that can be used to record holograms. This is by no means an exhaustive list. Many other materials have been used, at least experimentally. The table lists the resolution available and the energy per unit area required to record a hologram. These numbers should be regarded as approximate, because they will vary with the laser wavelength.

The materials listed in Table 18-5 employ a variety of physical or chemical effects to record holograms. The metallic bismuth, in thin film form, is simply vaporized

Table 18-5 **Selected Recording Media Used in Holography**

Material	Type	Sensitivity (erg/cm^2)	Resolution (lines/mm)
Dichromated gelatin	Phase	20,000–50,000	5000
Photopolymer (OmniDex 352)	Phase	10^5–10^6	—
Thermoplastic	Erasable phase	200	>1100
Metallic film (bismuth)	Amplitude	500,000	1000
Magnetic film (manganese bismuth)	Magnetooptic, erasable	300,000	500

in areas where the irradiance is high, to form an absorption hologram that needs no developing. The manganese bismuth films are heated and demagnetized in areas where the light is most intense; the reconstructed image is then formed through magnetooptic effects. Intense light causes the surface of thermoplastics to deform in the presence of an externally applied electric field, so as to form a hologram that needs no chemical developing. The action of short-wavelength light causes photopolymers to polymerize. These materials undergo a variation in index of refraction in regions where the light was most intense, so as to produce phase holograms.

Dichromated gelatin has been mentioned before as a recording material, and it is worth some discussion, because it is the medium that has given holograms with the highest efficiency. It has high resolution, because there are no silver grains to define a minimum feature size. Many striking types of holographic display have employed dichromated gelatin as the recording medium. The recording process involves tanning or crosslinking of the gelatin in regions of high light intensity, which in turn leads to variations in index of refraction. The chemistry of the photorecording process in dichromated gelatin is understood only poorly. Variations in the processing of dichromated gelatin have been discussed in the literature [2,3]. Dichromated gelatin is a difficult medium to control, and some irreproducibility and batch-to-batch variations are encountered. Also, the holograms must be sealed to protect them from moisture, which will degrade them over a period of time.

Because dichromated gelatin is unstable, it is not commercially available. It must be prepared by the user within a relatively short time (within one day) prior to use. It may be prepared by bleaching out the silver halide in the emulsion layer of a photographic plate. One such procedure that begins with 649F photographic plates is summarized in Table 18-6, along with the developing process [4].

The materials in Table 18-5 are used instead of photographic film when their unusual features offer an advantage. For example, thermoplastics and manganese bismuth offer erasability; dichromated gelatin offers very high efficiency. Some of the materials may be developed in situ, without wet chemical processing, which may be an advantage in some applications.

The end result of recording a hologram in any of these materials is a recording of a pattern of interference fringes in the recording medium. The method of forming the fringes varies and the result may be a hologram of any of several types (see Table 18-1) according to the way in which the hologram was made. The result is always the storage of the pattern of fringes caused by the interference of the object and reference beams. Just as with holograms formed in photographic film, the pattern has no obvious relation to the original object. But when the fringe pattern is reilluminated, it diffracts the light so as to form an image of the object.

The white light holograms that have become familiar in applications like displays, magazine covers, and credit cards are replicated in large quantity. The original master hologram is made with a laser. But to produce the large numbers of holograms required for a mass audience, a lower-cost nonlaser replication process is used.

Table 18-6 **Preparation and Developing of Dichromated Gelatin Plates, Derived from 649F Photographic Plates**

Procedure	Material	Time (min)	Lighting
Preparation of Plates (Temperature 20°C unless noted otherwise)			
Soak	Fixer without hardener	10	Normal
Wash, raising temperature slowly from 20 to 33°C, then lowering slowly to 20°C	Running water	5	Normal
Soak	Fixer with hardener	10	Normal
Wash	Running water	10	Normal
Rinse	Distilled water	5	Normal
Rinse	Photo-Flo solution	0.5	Normal
Drain, wipe, and dry			Normal
Soak	5% Ammonium dichromate solution with Photo-Flo	5	Dim red
Drain, wipe, and dry in horizontal position			Dim red
Development of Plates (Temperature 20°C unless noted otherwise)			
Soak	0.5% Ammonium dichromate solution	5	Dim red
Soak	Fixer with hardener	5	Dim red
Wash	Running water	10	Normal
Rinse	Photo-Flo solution	0.5	Normal
Soak	Distilled water	2	Normal
Dehydrate	Distilled water and isopropanol solution (50/50)	3	Normal
Dehydrate	Isopropanol	3	Normal
Dry		60	Normal
Bake or vacuum dry		120	Normal

In one process, which has been used for magazine covers, a white light hologram is recorded in photoresist, a photosensitive medium that, after processing, stores the fringes as a surface relief pattern. Then, nickel is deposited by electrolysis onto the surface so as to form a mold. The nickel mold then is used to impress the pattern onto plastic. In this way, many copies of the surface relief pattern that represents the hologram may be fabricated inexpensively. The plastic has a thin aluminum coating deposited and then is sealed and mounted on the magazine cover to act as a white light reflection hologram. Holograms produced in this way or in processes similar to this are called embossed holograms and may be mass produced at low cost.

We are now in a position to describe applications of holography, the subject of the next chapter.

References

[1] A. L. Rosen, *Appl. Phys. Lett.* **9**, 337 (1966).
[2] R. G. Brandes, E. E. Francois, and T. A. Shankoff, *Appl. Opt.* **8**, 2346 (1969).
[3] L. H. Lin, *Appl. Opt.* **8**, 963 (1969).
[4] B. J. Chang and C. D. Leonard, *Appl. Opt.* **18**, 2407 (1979).

Selected Additional References

A. Formation of Holograms

R. J. Collier, C. B. Burckhardt, and L. H. Lin, *Optical Holography*, Academic Press, New York, 1971, Chapter 1.
D. Gabor, Holography, 1948–1971, *Proc. IEEE* **60**, 655 (1972).
E. N. Leith and J. Upatnieks, Photography by Laser, *Sci. Am.*, p. 24 (June 1965).

B. The Holographic Process

H. J. Caulfield, *Handbook of Optical Holography*, Academic Press, New York, 1979.
R. J. Collier, C. B. Burckhardt, and L. H. Lin, *Optical Holography*, Academic Press, New York, 1971, Chapter 2.
J. B. DeVelis and G. O. Reynolds, *Theory and Applications of Holography*, Addison-Wesley, Reading, MA, 1967, Chapters 2 and 3.
H. M. Smith, *Principles of Holography*, Wiley (Interscience), New York, 1969, Chapters 2 and 3.

C. Hologram Types and Efficiency

R. J. Collier, C. B. Burckhardt, and L. H. Lin, *Optical Holography*, Academic Press, New York, 1971, Chapter 9.
E. N. Leith, White-Light Holograms, *Sci. Am.*, p. 80 (October, 1976).
T. A. Shankoff, Phase Holograms in Dichromated Gelatin, *Appl. Opt.* **7**, 2101 (1968).
H. M. Smith, *Principles of Holography*, Wiley (Interscience), New York, 1969, Chapters 3 and 4.

D. Practical Aspects of Holography

S. A. Benton, ed., *Practical Holography IV*, SPIE Proc., Vol. 1212, SPIE, Bellingham, WA, 1990.
R. J. Collier, C. B. Burckhardt, and L. H. Lin, *Optical Holography*, Academic Press, New York, 1971, Chapters 7 and 10..
P. Hariharan, *Optical Holography: Principles, Techniques and Applications*, Cambridge University Press, New York, 1986.
H. M. Smith, ed., *Holographic Recording Materials*, Springer-Verlag, Berlin and New York, 1977.

Chapter 19 | Applications of Holography

In this chapter, we will review some of the practical applications of holography. The most important application for industrial purposes is holographic interferometry, which can provide a valuable engineering tool for strain and vibration analysis and for defect detection. Thus, most of this chapter will be devoted to holographic interferometry.

It is also instructive to review, at least superficially, the wide range of other applications of holography. Some of these (like holographic movies) are still in an early stage of development. Other uses (such as an art form) are outside the scope of this book. Still, the variety of applications in this introductory survey can be stimulating to the imagination.

After completing the discussion of holographic interferometry, we shall describe briefly some other applications that can have an impact in industry.

A. Holographic Interferometry

Holographic interferometry is the area in which the most significant industrial applications of holography have developed. Conventional interferometry has commonly been used to determine surface contours for surfaces with relatively simple profiles. One example is the inspection of mirror surfaces and optical flats in the optics industry. Such interferometric measurements had previously been restricted to reflecting surfaces with simple shapes. This restriction is removed by the use of holographic interferometry, which makes interferometric methods applicable to the measurement and testing of surfaces with complicated shapes and to surfaces that are not specularly reflecting.

A hologram can be considered as a device that stores a wavefront representing an image of some object. The stored wavefront is released by the reconstruction of the hologram. In holographic interferometry, the wavefront is released and used to interfere with some other wavefront, so as to form bright and dark fringes in regions of constructive and destructive interference. The two wavefronts can represent an

object at different instants of time. For example, one may make a hologram of an object and then change the object slightly. If the reconstructed image of the original object, released from the hologram during reconstruction, interferes with a wavefront that represents the altered object, the interference pattern can yield sensitive information about how much the object has changed. Because one fringe will be formed when the object has its dimension changed by one wavelength of light, the sensitivity of this measuring technique is high, comparable to the wavelength of the light being used.

There are several ways in which the basic technique of holographic interferometry has been implemented, depending on how the different wavefronts representing the object have been obtained. We shall describe three of the most important types of holographic interferometry. These are

- Real-time holographic interferometry
- Double-exposure holographic interferometry
- Time-average holographic interferometry

After the description of these three approaches to holographic interferometry, we shall describe the interpretation of the fringe patterns so as to derive information on the deformation of objects.

1. REAL-TIME HOLOGRAPHIC INTERFEROMETRY

This method involves making a hologram of an object, altering the object, and allowing the light from the reconstructed image stored in the hologram to interfere with light from the altered object. Because the alterations in the object may be continuously varied, the interference pattern will change continuously. The changes in the object may then be monitored continuously by observing the motion of the interference fringes.

The steps in the procedure are as follows. First, one makes a hologram of the desired object. The hologram is developed using conventional photographic techniques and then is replaced in the position it occupied when the holographic exposure was made. The holographic image is reconstructed using the original reference beam. Thus, one obtains an image of the original object. This image is superimposed on the real physical object, which remains in position. One has two views of the object, one from the object itself and one from the reconstructed holographic image.

The two wavefronts representing these two images can interfere. If the path difference for light coming from the same point in the two images differs by one wavelength, one interference fringe will appear. Thus, if the object has changed since the hologram was made, interference fringes will be visible across the superimposed images, with each fringe representing a cumulative deformation of the object by one wavelength of light (or by one-half wavelength, depending on the geometry).

If the object has not changed in any way since the hologram was made, the two wavefronts will exactly coincide, and there will be no interference fringes. This situation is difficult to achieve in practice. It is hard to avoid distortion of the hologram. During development, photographic film may change dimensions somewhat. Also, it

is difficult to replace the hologram exactly in the position that it occupied during the original exposure. Development of the film in situ can eliminate the repositioning problem, but there usually will be some background fringes present even if the object has not changed. Careful technique can minimize the number of such fringes.

Let us consider an example. If the object is placed under stress and made to deform, the optical distance from the observer's eye to any point of the displaced surface will change. Light coming to the observer's eye from the two corresponding points on the image and on the deformed object will have a relative phase shift, which produces the fringe pattern. The fringe pattern defines the amount of deformation of the surface. If there are constant features on the surface, these will not give rise to fringes. Holographic interferometry will make small changes in the object stand out.

The motion of the fringe pattern can be seen immediately as the object is deformed. The fringe pattern will change and spread across the surface of the combined images according to the deformation of the object in real time, and changes in the object may be monitored continuously as they occur. Hence, the name real-time holographic interferometry is used to describe the process. The method is sometimes called live fringe holographic interferometry.

A schematic diagram of the operation of real-time holographic interferometry is shown in Figure 19-1. In the top portion of the figure, a hologram is made using holographic techniques as described earlier. In the middle portion, the hologram has been developed and replaced in its original position. The object, shown here as a can, has its end deformed by increasing the internal air pressure, so that the end bulges out. The result will be a circular pattern of fringes as shown in the bottom portion. The maximum displacement of the end of the can is at its center, so that the cumulative number of fringes will be largest at the center. If one were to observe this pattern develop in real time, one would see new fringes forming at the center and moving outward as the air pressure is increased.

The spacing and motion of the fringes can be measured and related to the amount of deformation of the surface. Thus, real-time holographic interferometry can provide a sensitive method for measuring strain of objects as they deform in real time. For example, if there is a weak area in the surface, it will deform more. The fringes will be crowded closer together around the weak area. Thus, this method may be used for detecting defects or other weakened regions in a structure.

One problem with real-time holographic interferometry involves the replacement of the hologram in exactly the position that it occupied during the original exposure. It must be repositioned accurately, within a fraction of the wavelength of light. To achieve this, the hologram mount may be equipped with micrometer drives for precise alignment. The positioning is accomplished by viewing the fringe pattern and varying the position of the hologram until the number of fringes across the image is at a minimum. This can be a time-consuming procedure. Alternatively, the hologram may be developed in place, so that there is no need for repositioning. Development systems have been produced that allow adding chemicals and processing the photographic film in situ, without any need for moving it. Also, use of the thermoplastic

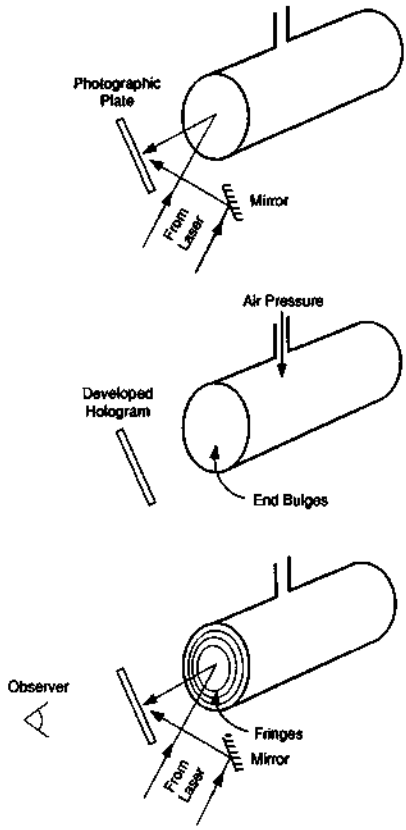


Figure 19-1 Holographic interferometry. In the top portion, a hologram is made of the end surface of a can. In the middle portion, the hologram has been developed and replaced in position. Increased air pressure in the can causes the end to bulge. In the bottom portion, interference between the two scenes (the physical object and the reconstructed holographic image) leads to a pattern of circular fringes.

recording medium described in the last chapter eliminates the problem of repositioning of the film. The thermoplastic is developed in situ through an electrical process without the use of wet chemicals. It may be erased by heat so as to record a succession of holographic interferograms. Commercial models of holographic recorders utilizing thermoplastic recording media and intended for industrial holographic interferometry are available.

As a very simple example of the use of holographic interferometry to determine a quantitative measure of a deformation, consider a horizontal beam, with its long axis in the x direction. The beam is fixed at one end, and its free end is displaced slightly in the vertical, or y, direction, so that the displacement is described by a function $Y(x)$. If a holographic interferogram of the beam is viewed from above, the

pattern of fringes will be a set of parallel lines perpendicular to the long axis of the beam. The phase shift $\Delta\phi(x)$ between the two wavefronts representing the beam and its holographic image is

$$\Delta\phi(x) = \frac{2\pi}{\lambda} 2Y(x) \tag{19.1}$$

where λ is the wavelength of the laser light. If one counts the bright fringes and numbers them $n = 0, 1, 2, 3, \ldots$ starting at the fixed end of the beam, then

$$Y(x) = \frac{n\lambda}{2} \tag{19.2}$$

Thus, one may determine the displacement easily simply by counting the number of fringes. We note the sensitivity of the approach; the displacement is readily determined to a fraction of the wavelength of light.

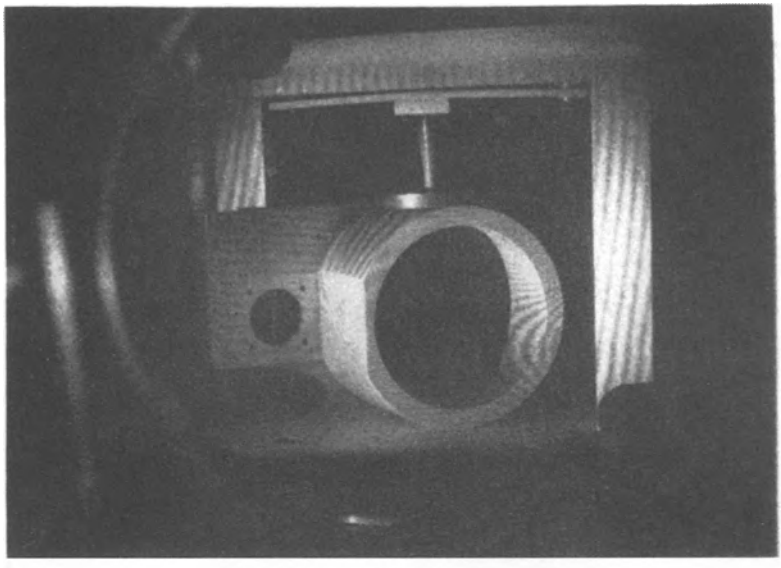
In general, however, it is much more difficult to interpret the fringe pattern so as to obtain quantitative measurements of the change in the object. If one desires information such as the presence of defects, the crowding together of fringes near the defect will reveal the presence of the defect, without the need for quantitative processing of the data. But when quantitative descriptions of the changes in the object are desired, data reduction can be tedious. We shall discuss the interpretation of the fringe pattern in more detail in a later subsection.

Figure 19-2 shows photographs of the fringe pattern formed in real-time holographic interferometry for an object with a relatively complicated shape. The hologram was made with no stress on the object. When stress was added, fringes were formed and the motion of the fringes could be observed as they moved in response to increasing deformation. Each fringe can be regarded as representing a constant change in the surface, so that the holographic interferogram is similar to a contour map. In this example, the deformation of the object could be monitored continuously as the stress increased. It is clear both that the interferometric technique dramatically shows the pattern of strain and that the mathematical reduction of the data to provide quantitative results will be complex.

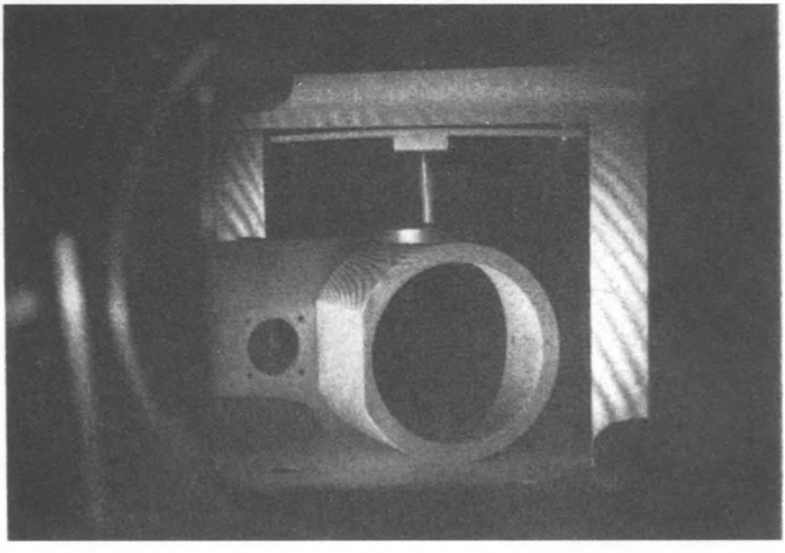
Real-time holographic interferometry has been used in a wide variety of applications, including areas like analysis of the strain of complex objects, observation of crack propagation, and determination of shock wave propagation.

2. *DOUBLE-EXPOSURE HOLOGRAPHIC INTERFEROMETRY*

Double-exposure holographic interferometry differs from real-time holographic interferometry in that two holograms of the object are made at different instants of time. In the two holographic exposures, the object will be slightly different. Thus, the double-exposure method compares the object in two different conditions. Two separate holographic images are obtained when the hologram is reconstructed. There is interference between the wavefronts representing the two holographic images. In contrast to real-time holographic interferometry, the object need not be



(a)



(b)

Figure 19-2 Fringe pattern formed in real-time holographic interferometry as a complex object is stressed. The stress is larger in (b) than in (a). (From K. A. Haines and B. P. Hildebrand, *Appl. Opt.* **5**, 595 (1966).)

present during the reconstruction. The wavefront that represents the object in its original condition is stored in the hologram, along with the wavefront representing the altered state of the object. The formation of the fringe pattern by interference between the wavefronts representing the object in two different conditions is the same as for real-time holographic interferometry.

This method is simpler and easier to carry out than real-time holographic interferometry. The double-exposure method avoids the problem of realignment of the hologram, because both images are stored in the hologram. The image may be reconstructed without taking any particular pains for exact repositioning and realignment. Distortion due to emulsion shrinkage is minimized, because emulsion shrinkage is identical for both exposures.

The double-exposure method can compare the original object with only one altered state of the object. Thus, it is somewhat less versatile than real-time holographic interferometry and will yield less complete information about the continuous change of the object. In many practical cases, continuous monitoring of the object's deformation is not necessary. Recording of relative surface displacement at a fixed interval of time can be useful. It can be used to detect defects by noting positions where fringes crowd close together. Because of its inability to monitor changes continuously, the double-exposure method is sometimes called a frozen fringe method.

The same comments hold about difficulties in interpreting the fringe pattern to obtain quantitative information about the surface deformation. Analysis of the fringe pattern to obtain the exact deformation of a surface is very difficult in the general case.

Figure 19-3 illustrates how double-exposure holographic interferometry can be applied to defect detection. The example is relevant to testing of tires. One exposure is made with low air pressure in the tire. The second exposure is made after the air pressure has been increased. The figure shows two examples of the fringe patterns obtained when the double-exposure holograms were reconstructed. A weak area on the tire deforms more when stress is applied. Thus, fringes are crowded closer together in the weak areas. These regions where the tires had low strength possibly represent defects.

3. *TIME-AVERAGE HOLOGRAPHIC INTERFEROMETRY*

A general rule applicable to most holographic situations is that the object should remain stationary during the period of exposure of the hologram. This general rule is violated dramatically in the case of time-average holographic interferometry. Time-average holographic interferometry is often used to study vibrating surfaces. During the exposure, the object is moving continuously. The hologram that results when the surface is vibrating at high frequency can be considered as the limiting case of a large number of exposures for many different positions of the surface. The mathematics describing the situation are complex, and we shall not deal with them. A conceptualization of the result can be obtained by considering a simplified model. The exposure of the continuously vibrating surface is similar to a double-exposure

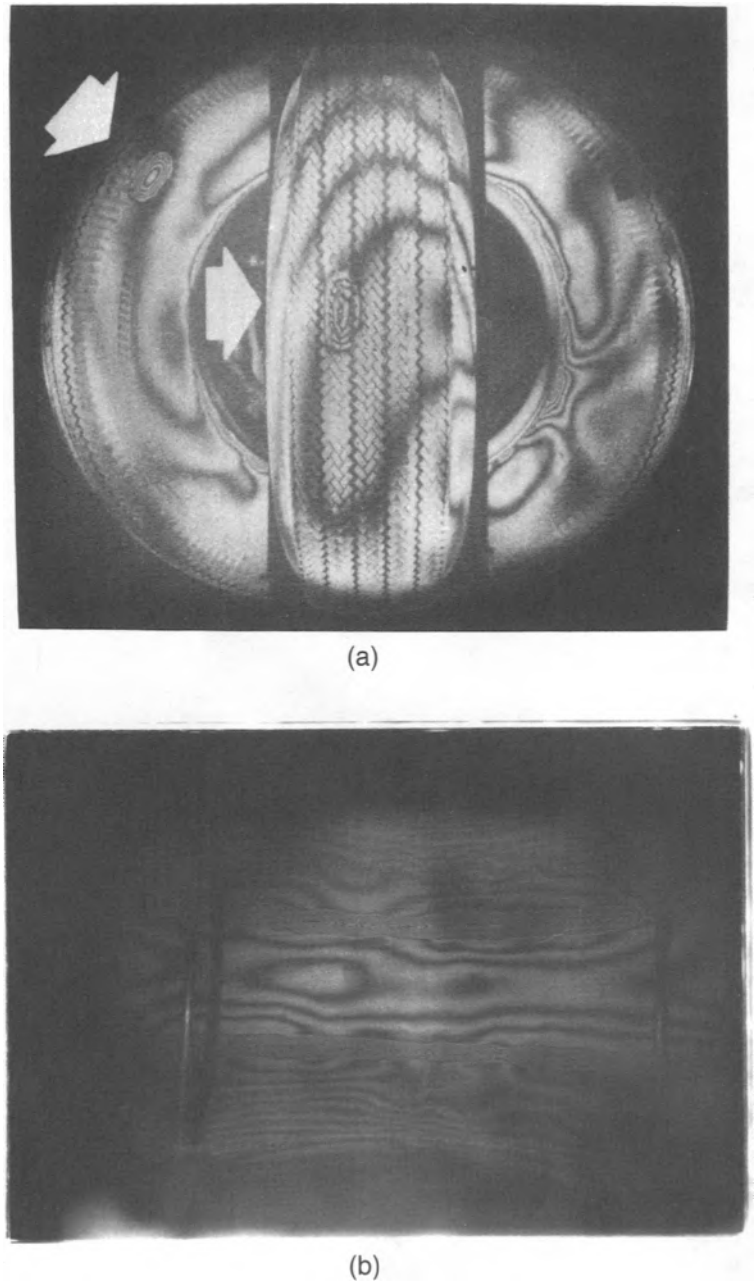


Figure 19-3 Double-exposure holographic interferograms of tires, in which air pressure was increased between exposures. In (a), arrows indicate areas of relatively large distortion, where fringes are crowded close together. These areas may represent defects. In (b), the areas where the fringes are close together show belt edge separation in each shoulder of the tire. (Photographs courtesy of GCO, Inc.)

holographic interferogram, in which the two different exposures represent the positions where the surface spends the most time. These two positions are the positions of extreme displacement of the surface. These are the positions where the motion stops and reverses direction. The speed of vibration is lowest in those regions. In this very simplified view, the two wavefronts are considered to originate from the surface at its extreme positions in its vibratory motion. The situation is sketched in Figure 19-4.

Time-average holographic interferometry is most useful in its application to vibrational analysis. With this method, one can measure the vibrational amplitude of diffusely reflecting surfaces with high precision. The method is simple to employ. It involves making a single hologram while the surface is vibrating. The period of exposure should be long compared with the vibrational period of the surface. The restriction, usual in holography, of no object motion during the exposure has been removed. The hologram is made just as if the surface had been motionless. After development and reconstruction, the resulting fringe pattern is measured to obtain information about the relative vibratory amplitude across the surface. Measurements of this type can be useful for determining modes of vibration of complex structures that are difficult to obtain using conventional techniques.

The reduction of the observed fringe pattern to obtain quantitative measurements of the amplitude of the surface vibration can be understood with reference to Figure 19-4. The light intensity $I(x)$ as a function of position x is given approximately by

$$I(x) = \frac{\lambda}{4\pi D(x)} \cos^2\left[\frac{4\pi D(x)}{\lambda} - \frac{\pi}{4}\right] \tag{19.3}$$

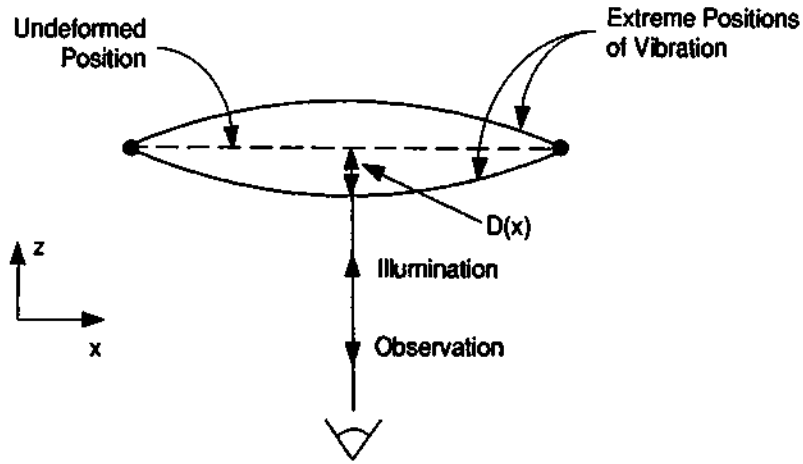


Figure 19-4 Schematic diagram for interpretation of fringes formed by a vibrating surface in time-average holographic interferometry. As one moves across the surface in the x direction, the maximum vibratory displacement of the surface from its undeformed position is $D(x)$.

where $D(x)$ is the displacement of the surface from its equilibrium or undeformed position, as a function of the transverse coordinate x. In this case, one fringe appears when the quantity $4\pi D/\lambda$ changes by π. We note that the change is not 2π, because the equation involves the square of a cosine function. Therefore, one fringe corresponds to a change of D by one-quarter wavelength of light. Counting fringes across the pattern thus yields the displacement as a function of position. This analysis is valid when

$$\frac{4\pi D}{\lambda} \gg 1 \tag{19.4}$$

Equation (19.3) breaks down when this inequality does not hold. In that case, more complicated mathematical expressions must be employed.

Equation (19.3) allows analysis of fringe patterns to obtain the maximum vibratory amplitude as a function of position across the surface. The amplitude of the fringes, according to Equation (19.3), decreases as a function of increasing displacement D. This decrease of intensity limits the number of fringes that can be observed. Thus, there is a maximum displacement of the surface from its equilibrium position for which time-average holographic interferometry will be useful.

A specific example of the application of time-average holographic interferometry in vibration analysis is shown in Figure 19-5. Vibrations of the bottom of a 35 mm film can were produced by a solenoid mounted inside the can and driven by an audio signal generator. In this figure, one type of vibration is observed when the can is excited at one particular audio frequency. Each closed fringe represents a contour of constant amplitude. Thus, the interferogram can be interpreted like a contour map. The vibration pattern is symmetric about a central line along which the amplitude is

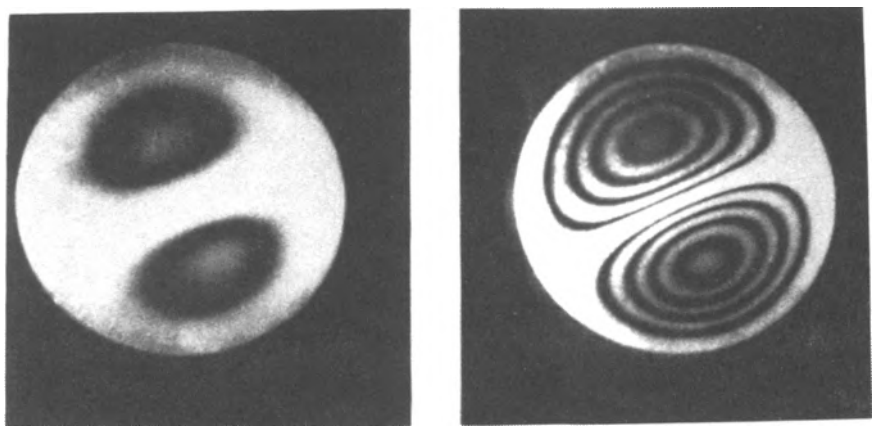


Figure 19-5 Reconstruction of the pattern of vibration of the bottom of a film can vibrating at an audio frequency. The fringes represent contours of equal vibratory amplitude. The frequency is the same in both photographs, and the amplitude of vibration is larger in the right photograph. (From R. L. Powell and K. A. Stetson, *J. Opt. Soc. Am.* **55**, 1593 (1965).)

zero. As one moves across the surface of the can in a direction perpendicular to this line, the vibratory amplitude increases from zero at the edge to a maximum, falls to zero at the central line, rises to another maximum, and falls to zero at the opposite edge. In the right portion of the figure, the voltage applied to the audio generator was increased above that for the left photograph. The vibratory amplitude was greater, and more fringes are visible. The exact values of the amplitude could be obtained by application of Equation (19.3). The figure illustrates the relative ease of analysis of vibratory motion by holographic techniques compared with that by other methods.

The technique of time-average holographic interferometry has proved useful for analysis of the motion of the surface of musical instruments, for analysis of turbine blade vibrations in the aircraft industry, and for study of vibrations of engines, frames, and brake systems in the automotive industry.

4. INTERPRETATION OF HOLOGRAPHIC INTERFEROGRAMS

In many cases of practical interest, information about the presence of defects may be gained by identification of areas in which fringes crowd close together, and it is not necessary to reduce the data to obtain quantitative estimates of deformation. If one seeks to obtain quantitative information about the deformation of an object, the analysis may become difficult.

The discussion in this subsection applies mainly to real-time and double-exposure holographic interferometry. As we discussed earlier, the interpretation of time-average holographic interferograms is relatively easy when the edges of the vibrating surface do not move. Then, one may interpret the result similarly to a contour map and simply count the number of fringes moving away from the stationary edge.

But in many cases the practical interpretation of a real-time or double-exposure hologram is difficult. Part of the problem is that the interference fringes do not lie on the object surface, and thus one cannot in general count the number of fringes starting at some known location. The interference fringes in holographic interferometry are localized on some surface. The surface of localization is a surface on which the fringes can be observed with highest contrast by an observer using an apertured optical system. The location of this surface will depend in a complicated fashion on the parameters of the experimental arrangement. But usually it will not be the surface of the object. This fact causes the difficulty. One cannot focus the optical system so as to view the object surface and the high-contrast fringes simultaneously.

We first consider a simple case, in which the surface of localization is the object surface. One such case in which the fringes are localized on the object surface is a pure rotation of the object about an axis that lies in the surface, with observation of the fringes in a direction perpendicular to the surface. The geometry is shown in Figure 19-6. The z axis is the direction of illumination and of observation. The phase shift δ between the two waves coming from the holographic image and from the original object is

$$\delta = \frac{4\pi x\theta}{\lambda} \tag{19.5}$$

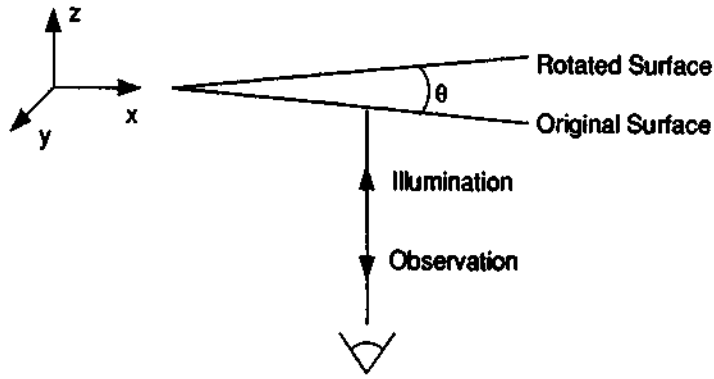


Figure 19-6 Schematic diagram for interpretation of a fringe pattern formed in holographic interferometry when a surface is rotated through an angle θ about an axis in the surface. The fringe pattern will consist of parallel fringes in the y direction.

where x is the distance along the x axis, as indicated, θ the angle of rotation, and λ is the wavelength of the light. The intensity I of the fringes is

$$I = C(1 + \cos \delta) \tag{19.6}$$

where C is approximately constant over the fringe pattern. Thus, an observer will see straight line fringes with intensity varying as a cosine function. If one moves a detector parallel to the surface in the x direction, one will obtain a signal that varies with a cosine dependence. The maximum intensity seen by the detector will occur when the phase shift is equal to an integral multiple of 2π. From Equation (19.5), the maxima will be separated by a distance equal to $\lambda/2\theta$. Measurement of the spatial frequency of the fringes thus easily yields the angle of rotation of the surface. This example shows how the surface deformation may be obtained quantitatively in one simple case.

In most cases, the situation is more complicated. If the deformation is a translation or a combination of translation and rotation, the fringes will be localized on some other surface in space, which may be anywhere between the original object surface and infinity. It is the failure of the fringes to be localized on the object surface that makes analysis difficult in the general case.

In general, the deformation of the surface at a particular point will be a combination of a translation and a rotation, and the magnitudes of the translation and rotation will vary from point to point. The fringes will not be localized on the object surface, and the position at which the fringes are localized will vary as the direction of observation changes.

A number of techniques have been developed by which surface displacement may be derived quantitatively from measurements of the fringes. Such procedures may be relatively complex, and we will not list them all here. We shall simply summarize one procedure.

The relation of surface displacement to fringe observation can be understood with reference to Figure 19-7, in which an imaging lens is focused on the object surface. The surface of fringe localization is different from the object surface. The lens has its axis oriented along the line PA, where P is a point in the surface of localization. The image plane of the lens is apertured so that one observes light originating only from a small area surrounding the point A on the object surface. The smaller the aperture becomes, the greater is the depth in which fringes may be viewed with good contrast. When the aperture is small enough, the fringes and the object may be viewed simultaneously.

The lens focuses through the hologram onto the original and deformed surfaces, which are nearly coincident, and simultaneously allows observation of high-contrast fringes. One then determines the displacement of the small area surrounding A on the surface. One observes the fringes that appear near A. Suppose that the observa-

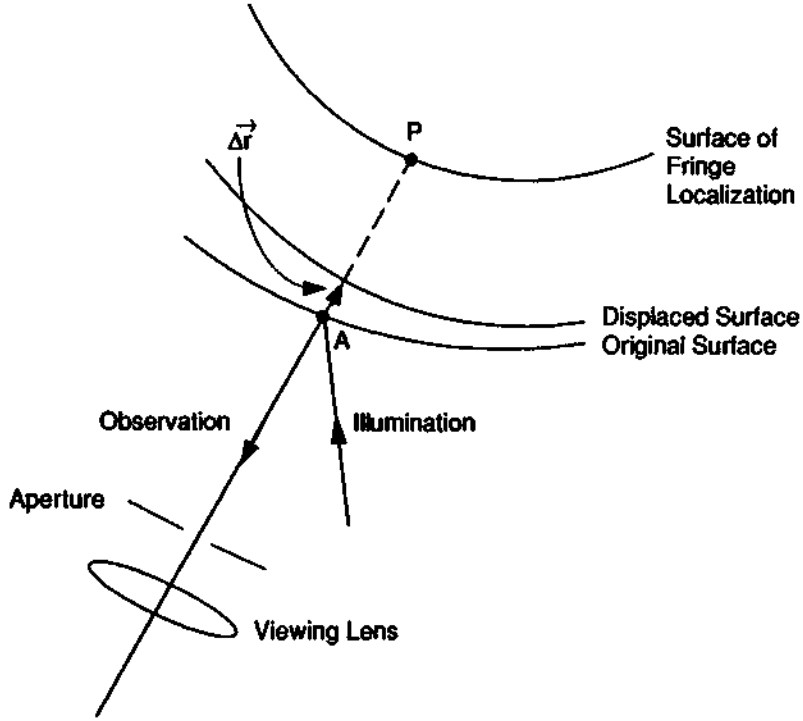


Figure 19-7 Schematic diagram for interpretation of fringe patterns in holographic interferometry when the fringes are localized at a position not on the object surface. The point A on the surface, which is being viewed, is displaced vectorially by $\Delta\mathbf{r}$. The axis of the viewing system intersects the surface of fringe localization at point P.

tion is made in a direction represented by a unit vector $\mathbf{n}_1$. The phase difference δ_1 between rays from the area A for the two surface positions is

$$\delta_1 = \frac{2\pi}{\lambda} \Delta\mathbf{r} \cdot (\mathbf{n}_0 - \mathbf{n}_1) \tag{19.7}$$

where $\Delta\mathbf{r}$ is the vector displacement between the two surfaces and $\mathbf{n}_0$ is a unit vector in the direction of illumination. Then, the area around A is viewed from a new viewing direction represented by the unit vector $\mathbf{n}_2$. As one changes viewing direction, the number of fringes m passing through A is counted. The direction of illumination is unchanged, and the phase difference δ_2 is

$$\delta_2 = \frac{2\pi}{\lambda} \Delta\mathbf{r} \cdot (\mathbf{n}_0 - \mathbf{n}_2) \tag{19.8}$$

Subtraction then yields

$$\delta_1 - \delta_2 = \frac{2\pi}{\lambda} \Delta\mathbf{r} \cdot (\mathbf{n}_2 - \mathbf{n}_1) = \pm 2\pi m \tag{19.9}$$

where $2\pi m$ is the number of radians representing the motion of m fringes through A. The appropriate sign must be determined from other considerations. This equation is a linear equation in the three unknown components of the displacement $\Delta\mathbf{r}$. If one then carries out two more pairs of measurements, one has two additional equations from which to determine all three components of the displacement.

This procedure will yield a quantitative specification of the displacement. But it is time consuming and leads to information about the displacement at only one point on the surface. A general characterization of the surface deformation would involve additional measurements at many other points.

Development of methods for determination of three-dimensional surface motion from the fringes obtained by holographic interferometry has been the subject of continuing research for years. Many methods have been proposed and demonstrated. But as the preceding example shows, they may be cumbersome and involve extensive measurement and calculation.

In one classification scheme [1], the various approaches to interpretation of holographic interferometry are grouped into four basic methods:

The fringe localization method This approach is based on the fact that the fringes are usually localized at a distance behind the image. This distance is determined, and the components in the plane perpendicular to the direction of observation are determined from measured fringe spacings.

The fringe counting method This method again uses the result that the fringes are localized at a distance from the image. One measures the parallax between the fringes and the image as the direction of view is changed.

The zero-order fringe method If one can identify some position in the image that has not changed, one can count the number of fringes from that position to obtain the displacement along the line of sight.

The hologram fringe method In this method, a small limiting aperture at the image gives a large depth of field. This allows fringes to be observed in any plane, in order to derive the three-dimensional displacement of any point on the object.

Table 19-1 compares the properties and limitations of these four approaches to quantitative interpretation of holographic interferograms. Reference [1] also defines the conditions for which each of these methods will be appropriate to use, and it offers guidelines to allow workers to choose an interpretation method that is advantageous for their particular application.

5. *SUMMARY OF HOLOGRAPHIC INTERFEROMETRY*

In the foregoing sections we have discussed some of the leading methods by which holography is applied to sensitive measurements of surface deformation. Our discussion has not covered the entire list of possibilities. It did survey some of the more important methods and illustrate the general principles. To compare these methods, we summarize the relative advantages and disadvantages and areas of application in Table 19-2.

In industrial applications, holographic interferometry is an engineering tool used for specialized measurements that are difficult to make by conventional procedures. It is employed by skilled personnel for applications like defect detection, surface contouring, studies of deformation of complex objects, vibration analysis, and studies of strain.

The application of holographic interferometry does have some limitations. It is an extremely sensitive technique, with potential resolution of the order of a fraction of the wavelength of light. Many practical situations do not require such high reso-

Table 19-1 **Summary of Procedures for Interpretation of Holographic Interferograms**[a]

Method	Fringe localization method	Fringe counting method	Zero-order fringe method	Hologram fringe method
Component of deformation	Perpendicular to line of sight	Perpendicular to line of sight	Along line of sight	All components measured
Views required for 1 component	1	Continuum	1	1 for each image point
Holograms required for 3D motion	1	1	3	1
Limitations	Must locate fringe plane accurately	Limited size of hologram limits accuracy	Must identify zero-order fringe	Time consuming
Accuracy	Very limited	Limited	Good	Fair

[a]Adapted from J. D. Briers, *Opt. and Quantum Electron.* **8**, 469 (1976).

Table 19-2 **Summary of Techniques for Holographic Interferometry**

Technique	Manner of implementation	Advantages	Limitations	Typical application
Real-time	Expose hologram, develop, replace, then stress object	Complete information about changes in object	Problems in interpretation, repositioning, emulsion shrinkage	Strain analysis, defect detection
Double-exposure	Expose hologram, stress object, reexpose hologram, then develop	Easy to make; no problems of repositioning or emulsion shrinkage	Less complete information than real-time, problems in interpretation	Strain analysis, defect detection, analysis of transient events
Time-average	Expose hologram while object is in motion, then develop	Easy to make, relatively easy to interpret	Not useful for static surfaces	Vibration analysis

lution. In these cases, the extreme sensitivity of holographic interferometry may actually be a disadvantage. If the deformation of a piece of material is large, holographic interferometry is probably not the technique of choice. In many practical industrial measurements, the motion of the piece extends over many thousands of wavelengths. The interference fringes would then be faint and difficult to count. This factor sets a practical limitation to the usefulness of holographic interferometry, to those situations where the motion is a relatively small number of optical wavelengths. In cases where the motion is larger, moiré techniques are more applicable [2].

We conclude this summary with two examples that illustrate the capabilities of holographic interferometry. Testing of laminates and composite structures can be performed easily by means of thermal stressing. If a hologram is made of the structure at ambient temperature and the interferogram is viewed in real time as the structure is heated slightly above ambient temperature, the fringes will be distorted as they move over disbonded areas. This procedure has been applied to a variety of laminated and composite structures, such as automotive clutch plates, simulated uranium fuel elements, graphite epoxy jet engine fan blades, and composite compressor blades. Figure 19-8 shows an example of disbonding in an automotive clutch plate, adhesively bonded to a steel substrate and thermally stressed by induction heating. Disbonded areas are easily detected.

Holographic interferometry has been used in the electronics industry to analyze the heating patterns that arise when electronic circuits are turned on. As the circuit heats, closely spaced fringes identify regions of large deformation, which can indicate areas of excessive heat or stress on component leads. The observed patterns often make it easy to identify mechanisms that could lead to a failure. Figure 19-9 shows a

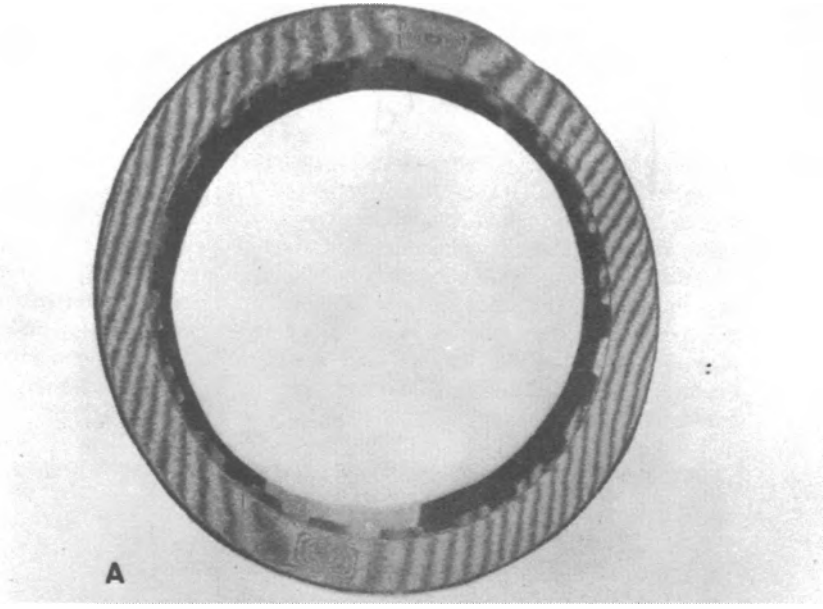


Figure 19-8 Double-exposure holographic interferogram showing areas of disbonding in an automotive clutch plate. The plate was stressed thermally between exposures. (From R. K. Erf, ed., *Holographic Nondestructive Testing*, Academic Press, New York, 1974).

holographic interferogram of a circuit board as it is heated. In the figure are indicated interpretations of the observed patterns and regions of potential failure.

B. Miscellany of Applications

No other application of holography is so well developed for industrial use as holographic interferometry. There are a number of other possible applications, some of which we will review briefly. This treatment will not be a complete survey, but it will serve to identify a variety of potential uses for holography.

1. MICROSCOPY

The technique of holography was originally suggested by Dennis Gabor for microscopy. The original purpose was to improve the resolution in electron microscopy. Gabor proposed a two-step holographic microscope in which the first stage would employ electron waves or x-rays to form the hologram and the second stage would use optical radiation to obtain a high-resolution reconstruction. The image obtained would be magnified approximately by the ratio of the wavelengths used in the two stages. Gabor demonstrated his concepts using optical radiation for both steps. The holographic concepts developed by Gabor in the late 1940s were

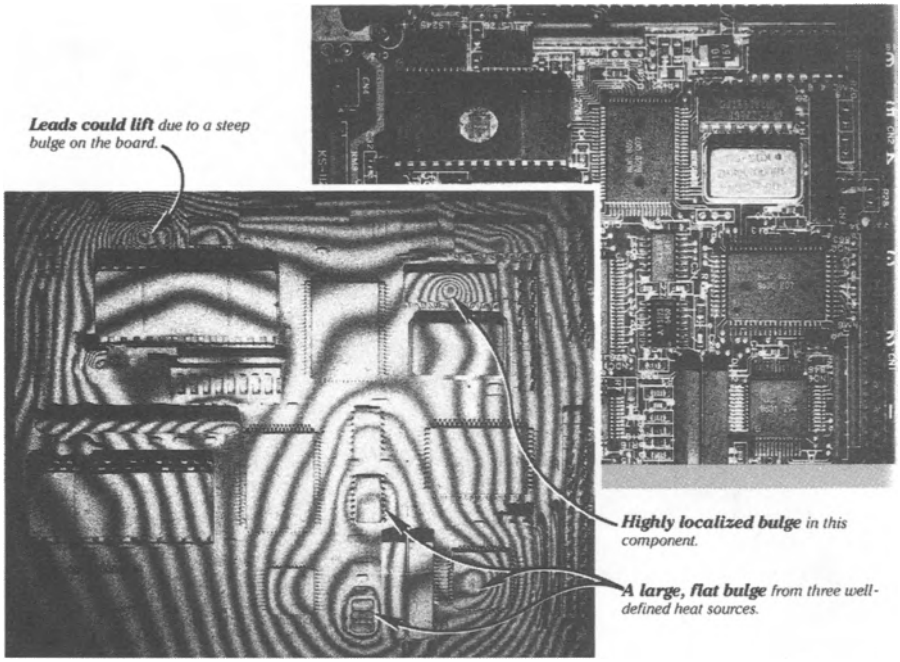


Figure 19-9 Holographic interferogram of a circuit board as it heats after being turned on. The arrows indicate regions of potential failure and causes for the observed patterns. (Photograph courtesy of Newport Corp.)

little used until the early 1960s, when the invention of the laser made available the coherent optical sources needed for high-quality holography.

The very high magnification that would result from the full implementation of Gabor's original idea of making holograms with electrons or x-rays and reconstructing them with visible light still lies in the future. Some examples of holography in the vacuum ultraviolet spectrum have been reported. For example, holograms have been made at a wavelength of 118.2 nm with the ninth harmonic of the Nd:YAG laser line at 1060 nm [3], using polymethyl methacrylate as a recording medium. Reconstruction of these holograms with a longer-wavelength laser operating at a wavelength of λ_v would yield a magnification equal to the ratio λ_v/118.2 nm. With visible lasers, magnification up to six would be possible. Gabor's concepts envisioned use of much shorter wavelengths for recording the holograms, so as to achieve much greater magnification.

The practical problems associated with achieving magnification comparable to that of an electron microscope are many, for example, lack of coherent x-ray sources and difficulty of control of holographic aberrations. These are some of the problems that must be solved to develop an x-ray holographic microscope with magnification extending possibly to atomic dimensions. Development of new, intense x-ray

sources could conceivably alter this situation and give to holographic microscopy the same impetus that the development of the laser gave to the field of holography as a whole. This is one reason for interest in the development of x-ray lasers. Recent progress in x-ray lasers has begun to be exploited for microscopic applications [4].

2. *FLOW VISUALIZATION*

Holographic techniques may be used to examine fluid flow patterns. Typically, double-exposure holograms are recorded with and without flow. Variation of refractive index due to the presence of the flow gives rise to optical path differences. The holographic interferogram preserves the optical path differences and gives an indication of the flow pattern. The resulting information is comparable to that yielded by conventional interferometry and by Schlieren photography. Because the information is obtained from one double-exposure hologram, the analysis may be easier than for the other techniques.

Flow visualization using pulsed lasers has been used to photograph transient shock waves. In one example, the pulse from a Q-switched laser was short enough to freeze the motion of a bullet. Figure 19-10 shows a double-exposure hologram, with the first exposure made of the undisturbed air and the second made with the bullet in the field of view. The discontinuity in the fringe pattern outlines the shock front produced by passage of the bullet. The change in density of the air and the position of the shock front can be derived from this reconstruction.

3. *HOLOGRAPHIC DATA STORAGE*

A hologram may be formed of a pattern of bits that represent data. The hologram can store this array of bits until they are needed. Then, the hologram can be reconstructed and the image projected onto an array of photodetectors to read out the stored information. Thus, a hologram can serve as the recording medium in a data storage system. Because an entire block containing a large number of bits is stored or read out at one time (instead of a single bit at a time), a holographic data storage system has the potential for very large capacity and for very high throughput rates.

Despite much early interest in holographic data storage, the practical problems have been formidable, and enthusiasm waned for a number of years. In the early 1990s, the development of new materials and new approaches has caused a resurgence of interest, and a number of research laboratories have resumed work in this field. The techniques will be described more fully in Chapter 24.

4. *ANTICOUNTERFEITING APPLICATIONS*

The use of holographic markings to curtail counterfeiting has become an application of major importance. These holograms are produced by the embossing process described in the last chapter. This process allows fabrication of large numbers of holograms inexpensively. It yields holograms that are contained in a metallized plastic matrix and that are incorporated into the packaging of a product in order to confirm

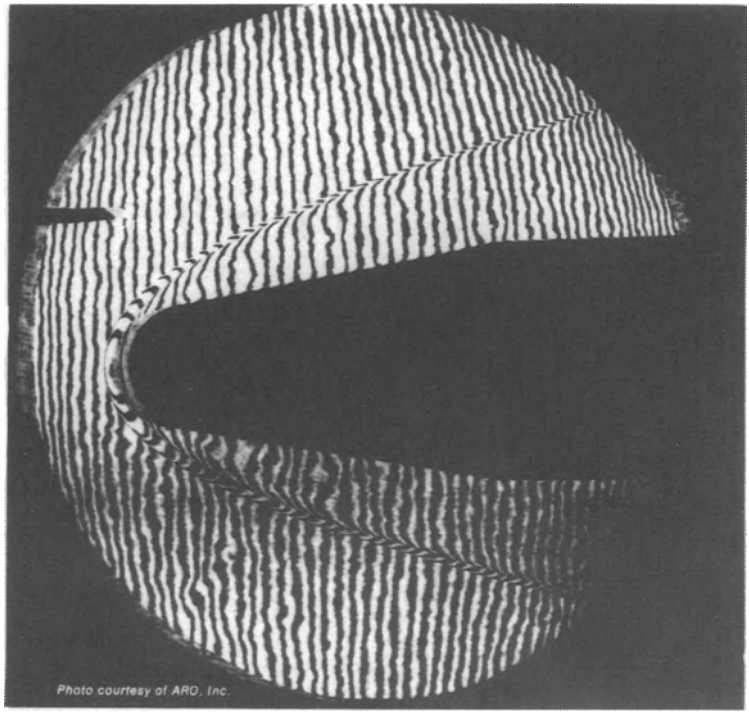


Figure 19-10 Double-exposure hologram of a bullet in flight, made with a Q-switched ruby laser. The feature of interest is the discontinuity in the fringe pattern, which is caused by the change in refractive index across the shock front produced by the bullet. (Photograph courtesy of ARO, Inc.)

it as genuine. Holographic anticounterfeiting markings have been incorporated into the packaging of many different types of products, especially for high-value brand names. Examples include videotapes, automotive parts, and perfumes. In addition, small embossed holograms have become a familiar feature on a number of different types of credit cards.

5. *IMAGING*

Holography has been employed in many applications involving imaging. We have already considered microscopy. Another well-developed imaging application has involved examining three-dimensional records in detail as a function of depth. Because the reconstructed image can be studied at leisure, the form and distribution of small objects can be determined fully. A hologram made with a pulsed laser freezes the motion of particles in the sample volume. The reconstructed three-dimensional sample volume may be examined plane by plane, and the details in each plane may be determined. Studies of distributions in this way by conventional photography are not practical.

This technique has been applied to the study of aerosols and the measurement of their particle distributions by optical sectioning of the holographic image. Current interest in knowledge of airborne particulates makes it important for manufacturers to be able to characterize aerosol products accurately.

In other investigations, microscopic flora and fauna have been examined through holographic micrographs to determine their distributions three-dimensionally at an instant of time in a way that was not possible before. Applications in nuclear physics for study of bubble chamber tracks have also been investigated.

The first direct use of holography in a practical application involved the sizing of small aerosol particles. Holograms of naturally occurring fog particles with a range of sizes from 4 to 200 μm were recorded. When the hologram was reconstructed, the image was magnified and displayed on a television screen. As the sample volume in the reconstructed image was scanned, different particles came into sharp focus at the appropriate locations in the image. This allowed the determination of particle size distributions within the fog as a function of depth.

6. *DISPLAYS*

Holography offers many possibilities for display because of the striking and attractive nature of the three-dimensional holographic image. White light holograms have been applied to advantage in a wide variety of different applications. One notable use has been the mass production of attractive magazine covers with embossed holograms. Embossed holograms are also used to enhance the attractiveness of paper and plastic products. Examples include holographic greeting cards and covers for compact discs and videotape boxes.

Holographic jewelry, including pendants, earrings, belt buckles, and so forth, has become commercially available. This application has usually involved white light holograms recorded in dichromated gelatin.

Holograms have been used in advertisements and in promotional materials. The use of holography for advertising purposes has developed only slowly, but this application seems certain to grow in the future. Holography has also developed as a separate art form, with many examples of strikingly beautiful artistry.

A number of novel techniques have been developed specifically for making holograms for display applications. Rainbow holography is one method that produces bright sharp images when the hologram is illuminated with white light [5]. The procedures for making rainbow holograms at first involved two steps. A procedure for making rainbow holograms in one step is illustrated in Figure 19-11.

The image from a rainbow hologram illuminated by a white light source is spectrally dispersed in the vertical direction. An observer whose eye is located at a particular vertical position will see a clear monochromatic image in the color at that particular height. Horizontal parallax is present when the observer's head is moved sideways. When it is moved vertically, there is no parallax, but the color of the scene changes.

Another display type that has become familiar is the white light holographic stereogram, in which the hologram is formed into a cylinder and illuminated with a

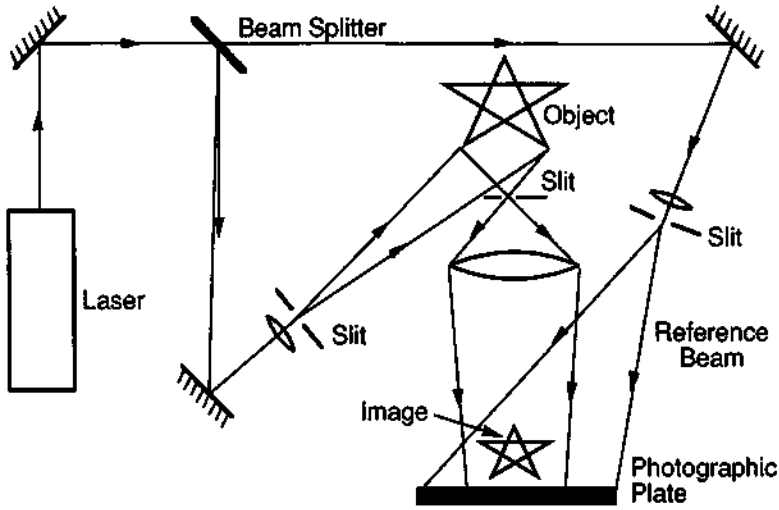


Figure 19-11 Optical arrangement for recording rainbow holograms in one step. (Adapted from F. T. S. Yu, A. M. Tai, and H. Chen, One-Step Rainbow Holography: Recent Development and Application, *Opt. Eng.* **19**, 666 (1980).)

white light source. As the observer looks around the cylinder (or as the cylinder is rotated about its central axis), some range of subject motion may be observed. A common approach to forming the holographic stereogram involves, as a first step, recording a series of ordinary two-dimensional photographs of the subject as the subject performs some action. For example, a movie camera may be used. Actions that have been recorded include such things as human beings dancing, blowing a kiss, and so forth. The original recording may include large scenes illuminated by natural light.

In the second step, transparencies of the series of photographs are projected onto a translucent screen, as illustrated in Figure 19-12. This arrangement records a series of holograms in the form of adjacent narrow vertical strips. As many as 1000 vertical strip holograms may be recorded. When the resulting set of holograms is viewed in white light, the observer sees a three-dimensional monochromatic image. If the holographic recording is formed into a cylinder and rotated, the observer sees the subject performing the set of actions originally photographed. Although there is a slight difference in the views seen by the observer's two eyes, the result is a convincing impression of three-dimensional motion within the cylinder.

7. *HOLOGRAPHIC MOVIES AND TELEVISION*

Startling applications in three-dimensional movies and television have been suggested. True holographic motion pictures (true in the sense that each frame is a hologram) have been generated experimentally. Such applications have involved the use of repetitively Q-switched lasers, operating at a pulse repetition rate compatible with

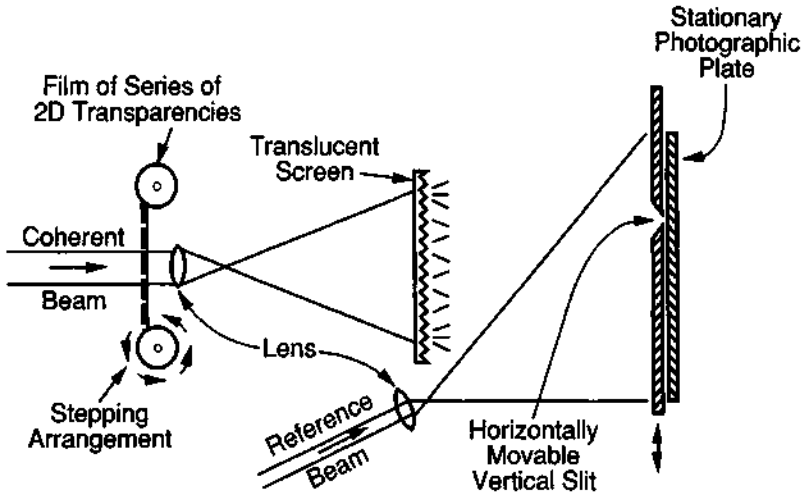


Figure 19-12 Top view of arrangement for recording a holographic stereogram from a series of two-dimensional transparencies. (From D. J. DeBitetto, *Appl. Opt.* 8, 1740 (1969).)

the motion picture framing rate. Each laser pulse records one frame of the motion picture. Such motion pictures have so far not been as pleasing artistically as conventional motion pictures. In addition, it has been difficult to produce holographic motion pictures that can be viewed by a large number of persons simultaneously. Holographic movies do not appear to be feasible as a large-scale entertainment form in the near future, although the long-term potential cannot be ruled out.

Similar remarks apply to holographic television. Three-dimensional holographic television has been demonstrated experimentally, but substantial advances in the formation of holograms, their transmission, and the viewing systems are needed before holographic television becomes practical.

Perhaps the best holographic representation of three-dimensional motion remains the holographic stereogram described in the last subsection. Although these stereograms can provide pleasing impressions of three-dimensional motion and can be viewed by a number of persons, the duration of the motion is limited to a few tens of seconds in current practice.

C. Holographic Optical Elements

Holographic optical elements (HOEs) are holograms that duplicate the performance of optical components, such as lenses. In this application, a hologram is used to transform the properties of a wavefront, just as some other optical component would. A holographic recording of a component like a lens will have the same optical properties as the component. It will focus light in the same way that the lens would.

Holographic optical elements have several attractive properties. They are very light in weight and can replace lenses in applications where weight is an issue. They can be replicated in quantity at relatively low cost. Because multiple holograms can be recorded in a single piece of recording material, it is possible to have multiple optical elements in the same plane. Perhaps the greatest disadvantage of holographic optical elements is that they work best for monochromatic light and it is difficult to design them for applications requiring broad spectral coverage.

Holographic optical elements have been used in a variety of applications, and their use is growing. They have been used to correct aberrations in optical systems. They have also been applied in heads-up displays for pilots, in which aircraft instrument readings are projected floating in space while the pilot retains a clear view of the space ahead of the airplane. Similar display systems have been proposed for automobiles.

Perhaps the most significant use of this technology has been in holographic scanners, which are employed in point-of-sale scanning systems. A number of holographic optical elements are attached to a rotating disk. The number may range up to 120, although around 20 is more common. The holograms deflect and focus the beam from a laser and scan the beam in a pattern, so as to cover the barcode on some item located in the image plane of the optical elements. Supermarket scanners employing holographic optical elements can be compact devices and offer the advantage that the object does not have to be drawn over the window in any exact manner.

D. An Example of Holographic Application

As a specific example, we shall describe the use of holographic interferometry to study thermal distortions of transparent materials. High-power lasers require window materials through which the laser beam can be transmitted. If any absorption is present in the window, it will heat and deform, producing distortion that can degrade the quality of the beam. Holographic interferometry has proved to be a valuable tool for studying such distortions, with the objective of reducing their effect on the beam. A description of this example will demonstrate the capabilities of holographic interferometry for solving difficult problems in materials research.

The holographic interferometer used to measure the distortion in materials for windows in high-power CO_2 lasers is shown in Figure 19-13. The light source is a He–Ne laser, the beam of which is divided by a beam splitter. Each resulting beam goes through an expanding telescope and a spatial filter. The slightly diverging reference beam goes directly to the recording medium. The object beam impinges on a diffuser plate, is transmitted through the sample, and reaches the recording medium, where it interferes with the reference beam and produces a hologram. The diffuser leads to a larger field of view for the reconstructed image.

When the hologram is illuminated by the reference beam, it reconstructs the wavefront characteristic of the object in its undeformed state. This wavefront then interferes with the wavefront from the sample itself. As the sample begins to deform

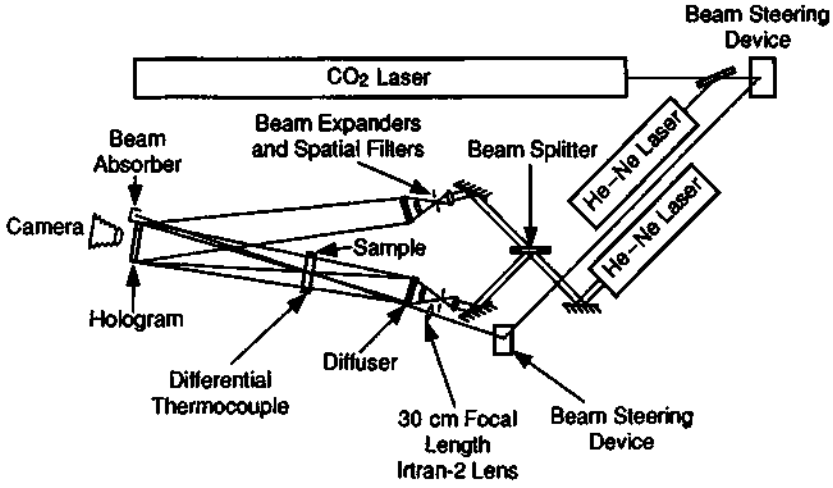


Figure 19-13 Holographic interferometer for measuring changes in optical thickness produced by CO_2 laser irradiation of transparent materials.

because of heating, interference fringes appear, each fringe corresponding to one-half wavelength of deformation. The holographic recording medium is an erasable thermoplastic material [6], such as has been described in Chapter 18. The thermoplastic is developed in situ, so that problems associated with plate repositioning are eliminated. Other advantages of the use of thermoplastics include reusability, low requirements for recording energy, high diffraction efficiency, and lack of grain noise.

The sample is irradiated by a beam from a 250-W continuous CO_2 laser, which passes through the center of the sample. The second He–Ne laser is boresighted with the CO_2 laser and is used for alignment. The CO_2 laser irradiance at the center of the sample could be as high as 13 kW/cm^2. Absorption of energy by the sample causes heating and optical distortion, which is recorded in the hologram.

Figure 19-14 shows one example of the sequences of interferograms photographed during irradiation of windows in the interferometer and during cooling after the laser is turned off. The sample is polished, uncoated polycrystalline potassium chloride with an absorption coefficient of 5×10^{-3} cm^{-1}, which is high for this type of material. The fringes are numbered so that the dark fringe nearest the center of the field of view in the undeformed state is zero. Fringes to the left have negative numbers; fringes to the right are positive. As time increases, and the sample heats, fringes initially in the field of view move out and new fringes move in, as indicated by the larger integers in some of the frames. The laser is turned off between the fourth and fifth frames. The sample then begins to cool, and fringes move back toward the right.

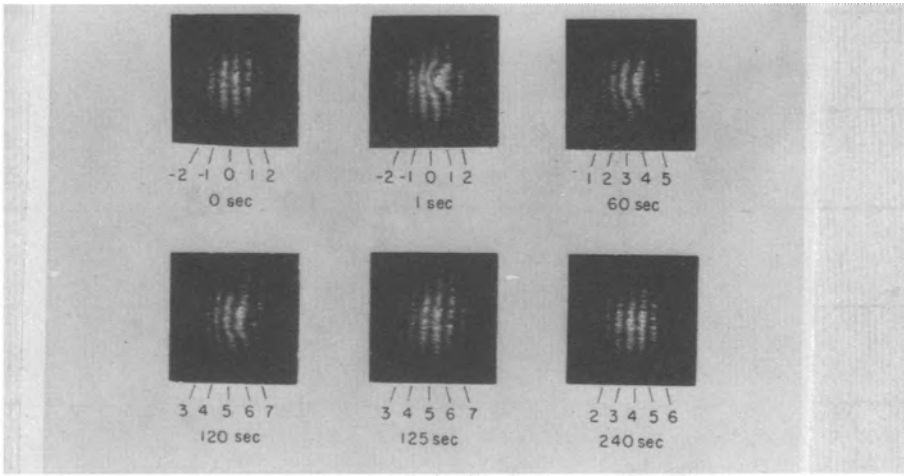


Figure 19-14 Interferograms showing wavefront distortion introduced by a high-absorption KCl sample during irradiation by a CO_2 laser beam delivering 13 kW/cm^2 near the center of the field of view, at the location of largest fringe curvature. The integers under each frame indicate the number of fringes from the center of the undeformed pattern, and the times are the times after CO_2 laser turn-on. (From E. Bernal G. and T. C. Lee, in *Digest of Papers, International Optical Computing Conference*, Washington, DC, April 23–25, 1975.)

Two qualitatively different effects can be seen in the interferograms. The first is the uniform motion of the fringe pattern across the sample, indicating a uniform temperature rise. The second is a localized deviation of the fringes from straight lines in the region where the beam strikes the sample. The uniform motion represents a relatively constant change in the optical thickness. It should have no effect on a collimated beam passing through the sample, although it would change the focus for a diverging or converging beam. The local deviations from straightness in the form of fringe curvature will have a serious effect on the quality of a beam passing through the window.

This example shows how holographic interferometry has been employed in a practical manner for solving a difficult problem in evaluation of optical materials.

References

[1] J. D. Briers, *Opt. and Quantum Electron.* **8**, 469 (1976).
[2] G. Oster and Y. Nishijima, *Sci. Am.*, p. 54 (May 1963).
[3] G. C. Bjorklund, S. E. Harris, and J. F. Young, *Appl. Phys. Lett.* **25**, 451 (1974).
[4] L. B. Da Silva *et al.*, *Opt. Lett.* **17**, 754 (1992).
[5] S. A. Benton, *J. Opt. Soc. Am.* **59**, 1545 (1969).
[6] T. C. Lee, *Appl. Opt.* **13**, 888 (1974).

Selected Additional References

A. Holographic Interferometry

T. E. Carlsson *et al.*, Practical System for Time-Resolved Holographic Interferometry, *Opt. Eng.* **30**, 1017 (1991).

H. J. Caulfield and S. Wu, *The Applications of Holography*, Wiley (Interscience), New York, 1970, Chapter XII.

R. J. Collier, C. B. Burckhardt, and L. H. Lin, *Optical Holography*, Academic Press, New York, 1971, Chapter 15.

P. Hariharan, *Optical Holography*, Cambridge University Press, Cambridge, UK, 1986, Chapters 14 and 15.

H. M. Smith, *Principles of Holography*, Wiley (Interscience), New York, 1969, Chapter 8.

A. Rebane and J. Feinberg, Time-Resolved Holography, *Nature* **351**, 378 (May 30, 1991).

C. M. Vest, *Holographic Interferometry*, Wiley, New York, 1979.

B. A Miscellany of Applications

S. A. Benton, Holographic Displays—a Review, *Opt. Eng.* **14**, 402 (1975).

S. A. Benton, Holographic Displays: 1975–1980, *Opt. Eng.* **19**, 686 (1980).

H. J. Caulfield and S. Wu, *The Applications of Holography*, Wiley (Interscience), New York, 1970, Chapters VIII–XI and XIII–XV.

R. J. Collier, C. B. Burckhardt, and L. H. Lin, *Optical Holography*, Academic Press, New York, 1971, Chapters 13, 14, and 16.

J. B. DeVelis and G. O. Reynolds, *Theory and Applications of Holography*, Addison-Wesley, Reading, MA, 1967, Chapter 8.

T. Inagaki, Hologram Lenses Lead to Compact Scanners, *IEEE Spectrum*, p. 39 (March 1989).

J. Robillard and H. J. Caulfield, eds., *Industrial Applications of Holography*, Oxford University Press, New York, 1990.

D. P. Towers, T. R. Judge, and P. J. Bryanston-Cross, Analysis of Holographic Fringe Data using the Dual Reference Approach, *Opt. Eng.* **30**, 452 (1991).

D. An Example of Holographic Application

J. S. Loomis and E. Bernal G., Optical Distortion by Laser-Heated Windows, in *Laser-Induced Damage in Optical Materials* (A. J. Glass and A. H. Guenther, eds.), National Bureau of Standards Spec. Publ. 435, U.S. Dept. of Commerce, Washington, DC, 1976, p. 126.

Chapter 20 | Laser Applications in Spectroscopy

The availability of laser sources has substantially changed the field of spectroscopy. Perhaps the most obvious form of laser-based spectroscopic measurement is simply to use a tunable laser, direct the beam through the sample to be measured, tune the laser wavelength, and measure the absorption spectrum of the sample. Thus, because of their high radiance and spectral purity, lasers have substantially advanced the capabilities of conventional spectroscopic techniques, like absorption spectroscopy.

But lasers have been employed in many more innovative approaches to spectroscopic investigation. Many new classes of spectroscopic methods and equipment have been developed to take advantage of the unusual properties of laser light. The laser is far more than simply a bright new source of tunable light. Rather, it offers new and highly versatile capabilities for novel measurements of the structure and dynamics of molecules. The availability of the laser has led to a qualitative revolution in spectroscopic capabilities and techniques.

Figure 20-1 illustrates the impact that lasers have had on the resolution of spectroscopic measurements. The top portion of the figure shows the absorption spectrum of sulfur hexafluoride at a wavelength near 10 μm obtained with a grating spectrograph. This represents capabilities available in the late 1960s with a resolution around 0.7 wavenumbers. (The wavenumber, expressed in reciprocal centimeters, is a measure of the photon energy associated with a particular wavelength.) The middle portion of the figure represents an expanded portion of the spectrum, obtained with a tunable diode laser. This represents capabilities available in the early 1970s, with a resolution around 0.001 wavenumbers. The bottom portion shows a further expanded portion obtained using saturation spectroscopy, to be discussed later. This represents capabilities available in the late 1970s, with a resolution around 10^{-6} wavenumbers. The figure illustrates the impact that lasers have had on spectroscopic capabilities.

In this chapter, we first discuss the types of lasers that are commonly used for spectroscopic studies, then we describe a few of the different types of laser spectroscopy that have been developed, and finally we describe some of the applications of laser spectroscopy.

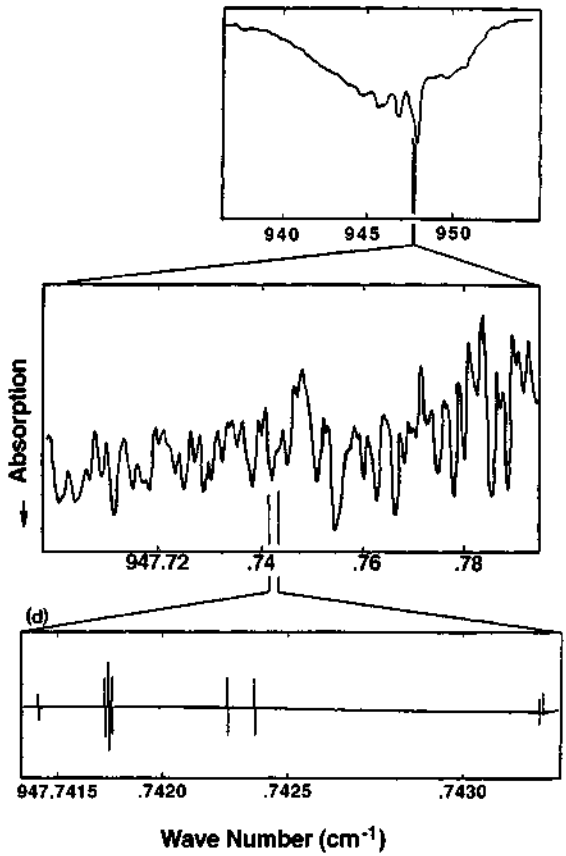


Figure 20-1 Absorption spectrum of sulfur hexafluoride near 10 μm. Top: Obtained with grating spectrograph. Middle: Obtained with tunable laser diode. Bottom: Obtained using saturation spectroscopy. (Based on information from L. J. Radziemski, Los Alamos National Laboratory.)

A. Lasers for Spectroscopic Applications

Lasers used in spectroscopy include many of the tunable lasers described in Chapter 4 and summarized in Table 4-2. The tunable dye laser has probably been the most commonly employed laser for spectroscopic applications. It has been used routinely in spectroscopic studies for many years. It is now being replaced in some applications by tunable solid state lasers, like Ti:sapphire, especially in the visible and near infrared regions.

Tunable lead salt lasers have been widely used for studies in the far infrared, and III–V compound diode lasers are beginning to be used in the near infrared. The tuning range of any specific device, though, is limited.

Other lasers that do not offer continuous tuning have also been employed in spectroscopic applications. The CO_2 laser, which offers line tuning over the range 9–11 μm, has been employed in far infrared applications.

Even lasers that strictly are not tunable, or at least are tunable to only a few discrete lines, have been used, especially for some of the more exotic types of spectroscopy, like Raman spectroscopy. The argon laser has frequently been used, both at its various fundamental wavelengths and as a frequency-doubled device with output in the ultraviolet. The Nd:YAG laser, as a frequency-doubled, -tripled, and -quadrupled device, offers several useful wavelengths, especially in the ultraviolet.

One limiting factor has been the availability of good laser sources in the ultraviolet, but advances in laser instrumentation, especially the use of nonlinear optics to generate ultraviolet sources, are removing this limitation.

B. Types of Laser Spectroscopy

A wide variety of configurations and methods for laser spectroscopy have been developed. Table 20-1 summarizes some of the many types of laser spectroscopy. The table lists the types, briefly describes their principles, and compares the materials for which they are useful, their applications, and their advantages and disadvantages.

Many articles in the technical literature describe these spectroscopic methods and their application. It is not possible to describe all the methods in detail in a reasonable space. We will consider a few selected types of laser spectroscopy as examples.

1. ABSORPTION SPECTROSCOPY

The first and perhaps most obvious type of laser spectroscopy is absorption spectroscopy, which involves tuning the laser. An example of the spectrum of sulfur hexafluoride in the 10 μm region, as obtained by tuning a semiconductor diode laser, has already been shown in the middle portion of Figure 20-1. Because of the narrow linewidth and high radiance of the laser, spectra with very high resolution may be obtained. The laser beam is transmitted through the sample the spectrum of which is desired. The intensity of the transmitted beam is monitored by a photodetector, and the wavelength of the laser light is tuned through the region of interest. There is no need for gratings, prisms, or any of the other dispersive elements used in conventional spectrometers.

The combination of narrow spectral linewidth, high radiance, and tunability available with lasers has led to great improvement in resolution for absorption spectroscopy. The resolution of a tunable laser system is far better than that of the best conventional dispersive spectrometers.

The status and availability of tunable laser sources suitable for applications in spectroscopy has been discussed in Chapter 4. Many types of tunable lasers are under development; these advances should continue to improve laser sources for spectroscopic applications.

Table 20-1 **Comparison of Laser-Based Spectroscopic Methods**

Type	Principle	Samples	Uses	Advantages	Disadvantages
Laser absorption spectroscopy	Resonant absorption of light	Mainly gases & liquids	Quantitative analysis, remote sensing	Species-specific, broad applicability	Wavelength & intensity stabilization needed
Laser-induced fluorescence	Emission of light from level populated by resonant absorption	Gases & liquids	Analytical spectrochemistry, species concentrations in flames	Selectivity, reduction of interference	Low laser output in UV, where many transitions are
Photoionization spectroscopy	Measurement of change in ionization equilibrium	Gases	Ultrasensitive detection	Very low detection limits	Costly & complicated
Photoacoustic spectroscopy	Absorbed light transformed to acoustic energy	Gases, liquids, solids	Analytic spectrochemistry	Simple & sensitive	Sample must be in contact with detector
Doppler-free spectroscopy	One beam saturates absorption, 2nd beam probes it	Gases	Very high resolution spectroscopy	Highest spectral resolution	Needs extra components
Laser-induced breakdown spectroscopy	Laser creates hot plasma, which emits light	Particles, surfaces, gases, liquids	Real-time spectrochemistry	Minimal sample preparation	Potentially difficult calibration
Raman spectroscopy	Excitation of molecular vibrational frequencies	Molecular gases, liquids, solids	Trace species identification	Species-specific, detects many species at same time	Background may cause high detection threshold
Coherent anti-Stokes Raman spectroscopy (CARS)	Two lasers coherently excite vibrational frequency	Molecular gases, liquids, solids	Trace species identification, composition profiling	Much higher signal than conventional Raman	Costly & complex
UV resonance Raman spectroscopy	Raman scattering from electronic state	Molecular gases, liquids, solids	Analytical chemistry, structure & dynamics of excited states	Highly versatile	Availability of suitable UV lasers
Differential absorption lidar	Absorption spectroscopy using distributed backscatter	Gases, aerosols, atmospheric particulates	Remote sensing	Allows monitoring of remote samples from one spot	May be low signal levels

Laser absorption spectroscopy has been carried out in the visible, near infrared, and near ultraviolet regions of the spectrum for many years using tunable dye lasers. More recently, Ti:sapphire lasers, including frequency-multiplied devices and covering most of the spectral range available to dye lasers, have been replacing dye lasers for many spectroscopic studies. Ti:sapphire and other tunable solid state lasers offer a number of advantages compared with liquid dye systems, and these lasers seem likely to become the leading source for spectroscopy in the visible, near infrared, and near ultraviolet regions.

Advances in tunable III–V compound semiconductor lasers have led to their use as spectroscopic sources in the near infrared (in the 1–2 μm region). In the longer wavelength infrared, lead salt semiconductor lasers have been used in spectroscopic applications for many years. These lasers are cryogenic and expensive. Further advances in laser technology are needed to provide sources for absorption spectroscopy in this region.

Many applications of laser absorption spectroscopy are available. They include as examples:

- Detection of trace components in automobile exhaust using far infrared tunable semiconductor diode lasers.
- Determination of the structure of atoms, molecules, and ions through measurement of hyperfine splittings, Zeeman splittings, and Stark splittings.
- Determination of relative populations of rotational and vibrational levels of molecules.
- Real-time measurements of the concentration and flow patterns of combustion products.
- Remote sensing of atmospheric pollutants. This application will be described in more detail in a later section.

2. *RAMAN SPECTROSCOPY*

The Raman effect involves scattering of light by molecules of gases, liquids, or solids. The Raman effect consists of the appearance of extra spectral lines near the wavelength of the incident light. The Raman lines in the scattered light are weaker than the light at the original wavelength. The Raman-shifted lines occur both at longer and shorter wavelengths than the original light; the lines at the shorter wavelengths are usually very weak.

The Raman spectrum is characteristic of the scattering molecule. The Raman lines occur at frequencies $\nu \pm \nu_k$, where ν is the original frequency and ν_k are the frequencies corresponding to quanta of molecular vibrations or rotations. If the scattered frequency is lower than the original frequency (longer wavelength), the incident light has excited a molecular vibration or rotation and the optical photon has decreased energy. This situation is called Stokes scattering. If the frequency of the scattered light is higher than the incident light (shorter wavelength), the light has gained energy from the vibrational or rotational quanta. This is called anti-Stokes

scattering. Because the values of ν_k are characteristic of individual molecules, investigation of the Raman spectrum provides a sensitive analytical tool.

Figure 20-2 shows a schematic example of a Raman scattering spectrum. A number of Stokes lines are present at frequencies lower than the original frequency, corresponding to absorption of one, two, three, and so forth, quanta of vibrational or rotational energy from the incident light. The spacings of the lines are equal and yield an identifying characteristic of the scattering molecule. A smaller number of anti-Stokes lines, with lower intensity, are present at frequencies higher than the original frequency. These lines correspond to transfer of one, two, and so forth, quanta of vibrational or rotational energy from the molecule to the optical field.

Raman spectroscopy has long been used for qualitative analysis and for identification of characteristic localized units of structure within molecules. The availability of laser sources has provided an important new device for use in Raman spectroscopy. The narrow linewidth and high radiance of laser sources make it much easier to identify the scattered light and to determine the amount of the wavelength shift. These same properties permit higher resolution of Raman spectra than was possible with conventional sources. The variety of available laser wavelengths makes it possible to carry out Raman spectroscopy while avoiding interfering absorption bands of the molecule being studied. With a tunable laser, the excitation frequency can be tuned to a resonant frequency of the molecule to produce a larger Raman signal. Although a tunable laser is not necessary for Raman spectroscopy,

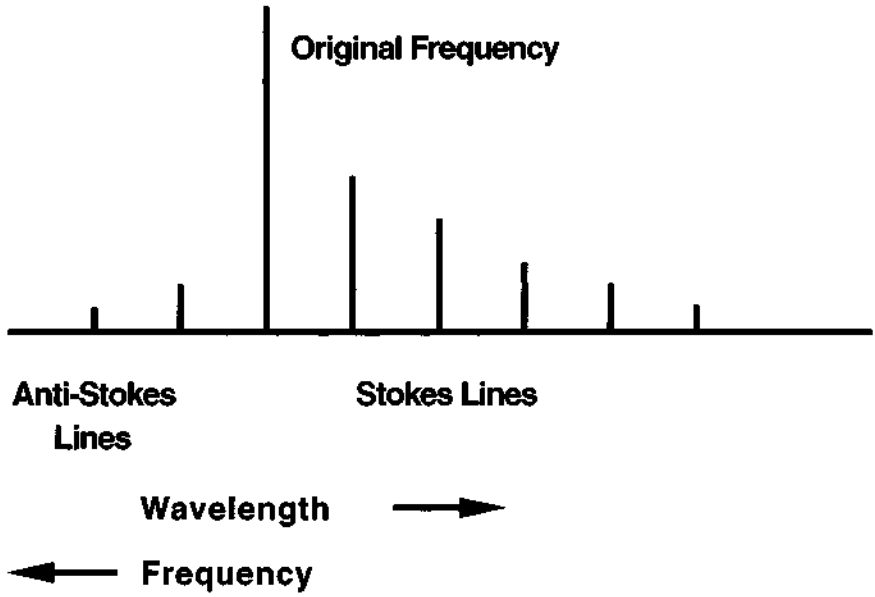


Figure 20-2 Schematic illustration of a Raman scattering spectrum. The separation between the different lines is equal to the frequency of a molecular vibration.

the use of a tunable source can greatly enhance the Raman signal. In practice, many of the studies involving Raman spectroscopy have been carried out with argon lasers.

Laser Raman spectroscopy has been used in a wide variety of applications, such as for identifying drugs, for detecting trace quantities of drugs in blood or urine samples, for detecting the metabolic by-products of drugs, for determining the presence of impurities in products like medicines, for studying polymers in solution, and for characterization of the kinetics of combustors and coal gasifiers.

Raman spectroscopy has also been used for remote detection of pollutants in the atmosphere. This application will be described in a later section.

The Raman spectroscopic technique is applicable to very small samples, because Raman spectroscopy requires a sample only of sufficient size to fill the focused beam of an argon laser, a volume about 10 μm in diameter. Thus, Raman spectroscopy can be useful for small samples of fibers, coatings, or finishes. Because Raman spectroscopy is a scattering technique, samples do not have to be transparent. Thus, materials such as pills, chemicals, drugs, small particles, and coatings can be sampled as received. These capabilities make the chemical laboratory an important market for laser Raman spectroscopic instrumentation.

One important variation of Raman spectroscopy is coherent anti-Stokes Raman spectroscopy (CARS). In CARS, one combines two laser beams with angular frequencies ω_1 and ω_2, within the sample to be measured. Then, one tunes ω_2 so that the frequency difference $\omega_1 - \omega_2$ is equal to the frequency of a Raman mode for the molecule of interest. Raman scattering then occurs through a nonlinear interaction at frequency $2\omega_1 - \omega_2$. The characteristics of the Raman signal yield a signature dependent on the molecular species, the temperature, and the pressure. Because of these capabilities, the CARS technique has been used for measurement of flow parameters. It is useful in environments with high-velocity gas flow and in situations where the gas is hot and luminous. The technique can provide both high temporal and spatial resolution.

3. *DOPPLER-FREE SPECTROSCOPY*

A limitation to resolution in absorption spectroscopy is the Doppler effect, the broadening of the spectral lines in gases and vapors because of the motion of the atoms or molecules. The spectral linewidths of absorption lines in gases should be very narrow. But the Doppler effect, caused by the random thermal motion of the molecules, can broaden the observed linewidth considerably. The Doppler linewidth, which is proportional to the square root of the temperature, can even be larger than the separation between neighboring narrow spectral lines. Thus, Doppler broadening can hide much of the fine structure contained in atomic or molecular spectra. The use of lasers in ordinary absorption spectroscopy does not help this limitation, because even though the laser linewidth is very narrow, the observed broadening arises from the motion of molecules in the sample. The broadening can be

eliminated by use of techniques termed Doppler-free spectroscopy. The resulting spectra reveal the natural width of the narrow spectral lines, not the width resulting from motion of the molecules.

One common Doppler-free technique is called saturation spectroscopy. In this approach, the laser beam is split into two beams, one strong and one weak. The two beams are introduced into a gas cell, traveling in opposite directions. The monochromatic laser beam, tuned to a specific wavelength, interacts only with a small fraction of the molecules. It interacts only with those molecules that happen to be moving with a velocity such that their absorption is at the same wavelength as the light.

The stronger beam can saturate the absorption, that is, it can interact with all the atoms or molecules in the narrow portion of the Doppler linewidth corresponding to the laser wavelength. Thus, the absorbing atoms or molecules are depleted, and the absorption is weakened in this portion of the absorption line. The stronger beam thus bleaches a path for the weaker probing beam. The strong beam is chopped with a modulator. If the two beams are interacting with the same molecules, then the probe beam will be modulated also. A detector viewing the probe beam sees the modulation. The modulated signal is detected only when the laser is tuned to the center of the absorption line, so that the light is absorbed by molecules having zero component of velocity in the direction of the laser beam. For those molecules, neither beam is Doppler shifted and the two beams are in resonance with the same molecules. In this way, the effects of Doppler broadening are eliminated.

Figure 20-1, in the bottom portion, has already shown how the resolution of a spectroscopic measurement may be increased by use of saturation spectroscopy. As another example, Figure 20-3 shows the hyperfine structure of a line of molecular iodine-127. The measurements were made with an argon laser that could be tuned over a very small range by a piezoelectric drive on one mirror. The Doppler width is shown for comparison. In conventional absorption spectroscopy, none of the detail on a scale narrower than the Doppler width would have been observable. The figure shows dramatically the ability of saturation spectroscopy to overcome the broadening introduced by the Doppler effect.

Another variation of Doppler-free spectroscopy is polarization spectroscopy, which developed later than saturation spectroscopy. Polarization spectroscopy yields greater sensitivity than saturation spectroscopy in the sense that it is possible to work with fewer molecules. The apparatus for polarization spectroscopy is similar to that for saturation spectroscopy, but the sample is between crossed polarizers, so that the background light is reduced to near zero. Because the probability for absorbing polarized light depends on the orientation of the absorbing molecule, the strong beam preferentially depletes the population of molecules with a specific orientation. Then, in its interactions with the oriented molecules, the weaker probe beam acquires a component of polarization that passes through the second polarizer. A detector then sees an increase of light intensity against a zero background. As with saturation spectroscopy, this method requires that both

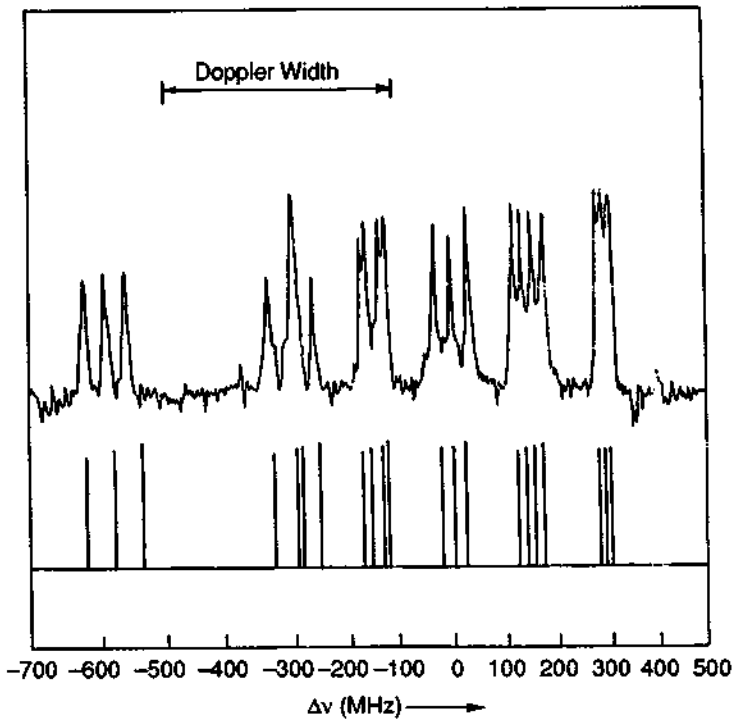


Figure 20-3 Hyperfine structure of an absorption line of $^{127}I_2$, probed by saturation spectroscopy. The frequency shift from line center is denoted $\Delta\nu$. (From A. L. Schawlow, in *Fundamental and Applied Laser Physics* (M. S. Feld, A. Javan, and N. A. Kurnit, eds.), Wiley, New York, 1973.)

beams interact with the same molecules near the center of the Doppler-broadened line.

C. Applications of Laser Spectroscopy

The unique properties of laser light, particularly its monochromaticity and high radiance, have made possible many advances in spectroscopic applications. Lasers provide greatly increased resolution in conventional spectroscopic measurements and also allow many innovative new spectroscopic techniques. Laser spectroscopy has become a very broad and varied subject, with many, diverse applications. It is not possible to describe all the many applications in chemistry, physics, medicine, life sciences, forensic science, and so forth that are made possible by lasers. We will simply present some examples.

1. CHEMICAL APPLICATIONS

Laser spectroscopy provides chemists with an extremely versatile tool for analytical spectrochemistry, for probing the dynamic behavior of chemical reactions, and for determining molecular structure and energy levels.

One significant application made possible by lasers is the technique of picosecond and subpicosecond spectroscopy, which is used to study ultrafast kinetics of chemical reactions. A number of tunable lasers, such as liquid dye lasers, can emit extremely short pulses with durations in the picosecond, or even femtosecond (10^{-15} sec), regimes. The availability of such lasers permits many new applications in the study of transient chemical phenomena. Examples include determination of the lifetimes of excited states and measurements of relaxation processes in atoms and molecules with extremely high temporal resolution.

Chemical kinetic processes can be monitored in a fashion that was not previously possible. Processes taking place on a picosecond time scale include vibrational and orientational relaxation of molecules in liquids, radiationless transitions of electronically excited large molecules, solvation of photoejected free electrons, and isomerization of photoexcited states of complex molecules like visual pigments.

We present one example that will illustrate some of the techniques and serve to demonstrate the capabilities of picosecond spectroscopy. The experiment [1] involved the formation and decay of a biophysical compound called prelumirhodopsin, which is considered to be the initial step in photoinduced chemical changes that begin the chain of physiological processes constituting vision. Vision begins when rhodopsin, a photosensitive pigment in the retina, is bleached by light. When rhodopsin in the eye is exposed to light, prelumirhodopsin is formed faster than can be followed by conventional methods.

The experiment involved photoexcitation of rhodopsin by a picosecond pulse of light at a wavelength of 530 nm. The absorption band of prelumirhodopsin at 560 nm was monitored by a train of short pulses at that wavelength. The original laser was a mode-locked Nd:glass laser, from which one pulse of 6 psec duration was extracted. This pulse was amplified and frequency-doubled in a potassium dihydrogen phosphate crystal to yield light at 530 nm wavelength. Part of that light was used to excite the rhodopsin. Another part was split off and Raman scattered to yield pulses of 560 nm light. This light was reflected from an echelon, a stepped optical element that introduced different optical delays into different spatial parts of the wavefront. Thus, the 560 nm light was transformed into a series of short interrogating pulses to monitor the desired absorption band. These pulses were focused onto the same area of the sample as the 530 nm exciting light. After emerging from the sample, the 560 nm light was imaged onto a camera. Light from each optical segment of the echelon was imaged on a different area of the film. Because the light from each element had a unique incremental optical path, the temporal pulse separation (about 20 psec) was recorded as a spatial separation.

The results of the measurement showed that the rise time of the prelumirhodopsin was less than 6 psec after the start of the 530 nm exciting pulse. The data indicated that production of prelumirhodopsin is the primary photochemical event.

Probing time-resolved chemical processes by means of picosecond spectroscopy has been the subject of extensive research, which has led to significant new knowledge about mechanisms of chemical reactions, energy distributions, and relaxation processes in molecules and structures of the excited states of biological systems.

2. *REMOTE SENSING AND ENVIRONMENTAL MONITORING*

There is a need for improved, accurate methods for remote measurement of the concentration of pollutants in the atmosphere. Many different types of pollutants are present in the atmosphere. Some of the major pollutants of interest include oxides of nitrogen, carbon monoxide, sulfur dioxide, and ozone. In addition, there are various types of particulate materials, such as ash, dust, and soot. There are also requirements to monitor sources of pollution, in order to comply with governmental regulations. Moreover, there are needs for mapping, on a global scale, constituents such as carbon dioxide and water vapor.

Many problems are associated with measurement of concentrations of pollutants in the atmosphere. Conventional measurements are difficult. For example, if one desires to measure gases from a smokestack, there are problems involving the inhospitable environment inside the smokestack, problems with accessibility, and problems involving perturbing the quantities that are to be measured. Traditional methods have involved collecting a sample of gas from the smokestack and making a chemical or spectroscopic analysis. Such methods are conducted remotely, and results are not immediately available. Stratification of gases and particulate materials also could lead to error.

Sensors embedded in the smokestack have used the electrochemical properties of the gases as a sensing mechanism. These sensors require contact with the gas sample and, for the most part, have slower response than desired. Such devices have had problems with maintenance because of the hot, corrosive gases that contact them. Although there have been numerous approaches to monitoring smokestack emissions, no completely satisfactory solutions have yet been developed.

The problem becomes even more severe when one considers monitoring the concentrations of specific gases remotely over broad areas. Approaches based on optical technology appear attractive for areal mapping. Fourier transform infrared spectroscopy (FTIR) based on nonlaser sources has been employed for this application. There are commercial FTIR monitoring instruments available.

Laser technology offers attractive features for remote monitoring of atmospheric quantities. It is possible to perform measurements at great distances. The measurements are completely noncontact and involve no sample collection nor chemical processing. There is no disturbance of the quantities to be measured. Generally, the results of the measurement can be available immediately.

Tunable semiconductor laser diodes with very narrow spectral linewidth offer many attractive opportunities for remote noncontact spectroscopic sensing of gaseous species. The use of laser spectroscopy for detection of pollutant gases has been studied for a number of years. The emphasis in early work has been on relatively long wavelength infrared spectroscopy, using lasers operating in the 5–10 μm

wavelength range. The lasers have been tunable lead salt diode lasers, such as lead selenide sulfide. Such lasers are expensive, operate at cryogenic temperatures, and require the use of expensive cryogenic detectors. Operation in the long wavelength infrared does offer the advantage that the molecular absorption lines are very strong for gases of interest.

A number of approaches for laser-based remote sensing have been studied over several decades. These include

- Optical radar (lidar)
- Raman backscattering
- Resonance fluorescence
- Absorption methods, particularly differential absorption laser backscatter (DIAL)

We shall discuss each of these approaches briefly.

1. Optical Radar

Optical radar, also called lidar, which stands for light detection and ranging, operates similarly to microwave radar. In its basic form, it employs a pulsed laser, the beam of which is directed to the atmospheric sample that is to be probed. Energy backscattered by the atmosphere is sensed by a photodetector. The resulting signal is processed as a function of time from the transmission of the pulse, just as in a radar ranging system. Generally, lidar systems have relied on Mie scattering, due to particulate material and aerosols, to provide the backscattered light. Thus, lidar is most useful for determining concentrations of particulate material, and it gives no information about gas concentrations. Lidar has most often been used for measurements in which concentrations of particulates or aerosols are desired. It can determine concentration as a function of distance from the measurement position.

Figure 20-4 presents what is perhaps a typical example of a lidar measurement. The results are derived from laser backscattering from the emission of an 800-foot-high smokestack. The figure shows the range-corrected signal in decibels relative to that from the ambient background aerosols. The contours represent relative particulate concentration in a vertical cross section of the emission from the smokestack. The example shows clearly how lidar can provide information about particulate concentration that would otherwise be difficult to obtain. The measurement may easily be performed remotely. Lidar can also measure quantities like stratification, flow, and changing profiles of turbid layers in the atmosphere.

2. Raman Backscattering

The use of the Raman effect, which involves scattering of light by molecules of gases with a shift in the wavelength of the scattered light, has been discussed previously in terms of its spectroscopic applications. It also has application to remote sensing, particularly for identification of gaseous species. Because the values of the

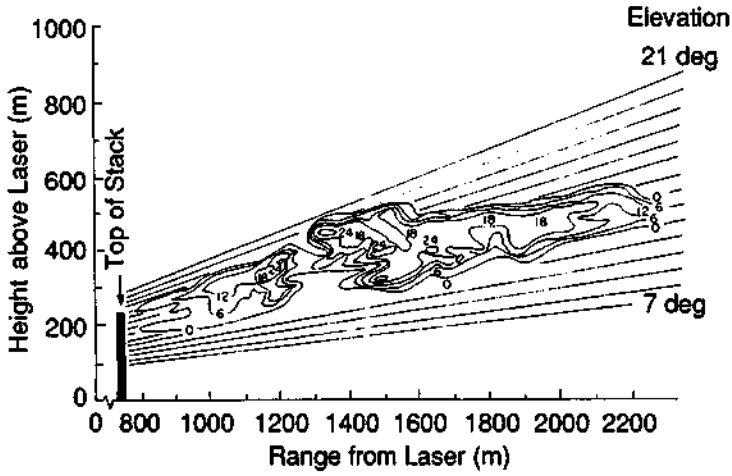


Figure 20-4 Vertical cross section through a smoke plume from a smokestack, as obtained by lidar measurements. The contours represent relative particulate concentrations. (From W. B. Johnson, *J. Air Pollution Contr. Assoc.* **19**, 176 (1969).)

wavelength shift are characteristic of individual molecules, examination of the spectrum of the light backscattered from a remote sample of gas can provide information about the gases present in the sample.

Figure 20-5 presents an example of the spectral composition of the backscattered return from an oil-smoke plume emitted by a smokestack at a distance around 30 m. The figure shows the strength of the return signal over a range of wavelengths. A number of pollutants, such as SO_2 and H_2S, are identified. Also, the major constituents of the atmosphere, oxygen and nitrogen, appear as relatively strong signals.

Raman-based remote sensing requires the use of visible or near ultraviolet lasers. Use of the Raman effect with a pulsed laser also offers range discrimination, so that relative pollutant concentration as a function of distance can be obtained. But because the cross section for Raman scattering is small, the laser used must have high values of peak power. This requirement leads to problems with eye safety. Also, Raman methods are mainly useful for measuring relatively high concentrations and at relatively short range. Although there have been many demonstrations of Raman-based remote sensing over a long period, the use of Raman sensing has not been adopted for routine monitoring use.

3. Resonance Fluorescence

In resonance fluorescence, the beam is directed to a gaseous sample that contains molecules that fluoresce when excited with the proper wavelength. An example is

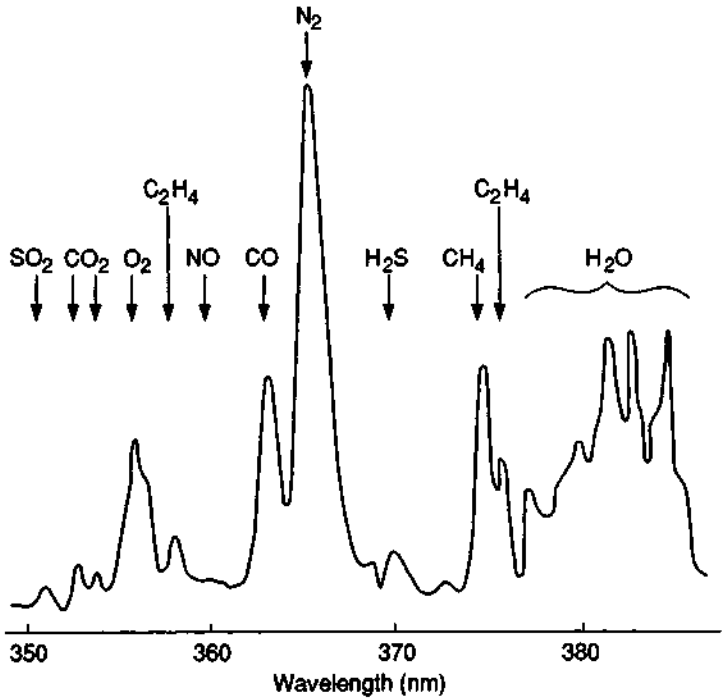


Figure 20-5 Spectral distribution of the Raman-shifted signal from a variety of molecular gases in a remote laser Raman measurement. (From H. Inaba and T. Kobayasi, *Opto-Electron.* **4**, 101 (1972).)

NO_2, which fluoresces when excited by blue light, such as 488 nm light from an argon laser. The fluorescence is collected by a telescope and delivered to a detector.

An example showing results of laser detection of NO_2, using an argon laser on samples collected in Los Angeles on a smoggy day, is presented in Figure 20-6. The concentration of NO_2 varies as a function of the time of day, with peaks near the times of peak automobile traffic. The rapidity of the changes shows the desirability of real-time monitoring with fast response.

In addition to use for atmospheric monitoring, fluorescence techniques have been used for measurements underground, to determine the presence of pollutants in groundwater and in sites where toxic wastes may be present. Test holes may be drilled, or a sensing element may be driven into the ground in a cone penetrometer. Such devices can be rammed tens of feet into the ground in suitable soils. A fiber optic cable is used to deliver laser light to the underground region being tested. Fluo-

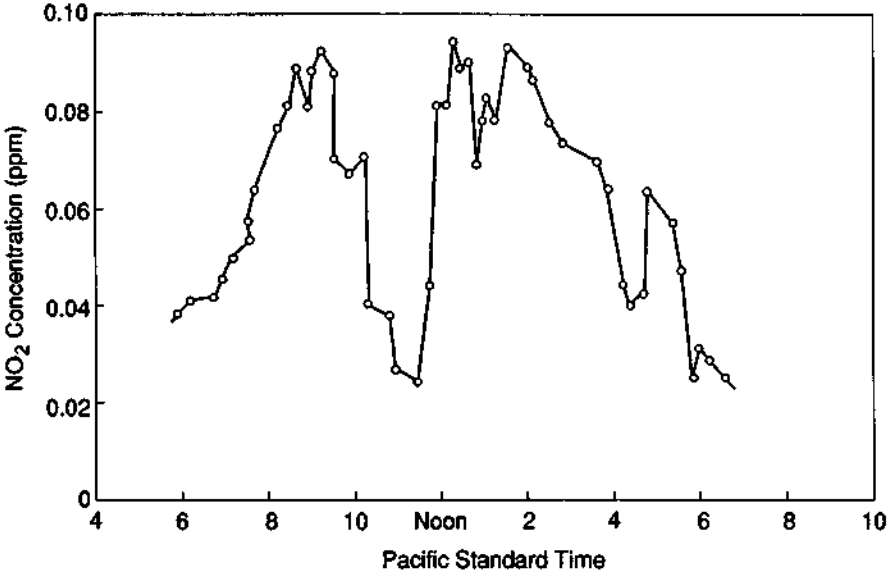


Figure 20-6 Nitrogen dioxide concentration measured as a function of local time during a smoggy period in the Los Angeles basin. The data were obtained using a resonance fluorescence technique. (From J. A. Gelbwachs *et al.*, *Opto-Electron.* 4, 155 (1972).)

rescence emitted by materials in the ground is collected and delivered back through the fiber optic cable to a detector at the surface. The presence of specific chemicals is revealed by the spectrum of the fluorescence.

In one example, an excimer laser was used with a Raman shifter to produce light at 248, 276, and 317 nm. The light was delivered through optical fibers into boreholes near a leaking gas pump. The fluorescence from hydrocarbons in the soil was returned to the surface by the fiber and detected with a monochromator and a photomultiplier. The presence of diesel fuel in the soil was detected, and the concentrations could be readily mapped. Figure 20-7 shows experimental results for diesel fuel contamination in a vertical layer of soil near the pump. The results demonstrate the usefulness of laser techniques for in situ monitoring of subsurface contaminants. The system could distinguish contaminants such as fuels, benzene, toluene, xylene, and polycyclic aromatic hydrocarbons.

The status of this type of measurement is still developmental, but a number of demonstrations have been performed. The technique offers the advantage of real-time in situ determination of the presence of underground pollutants, without the need for chemical processing.

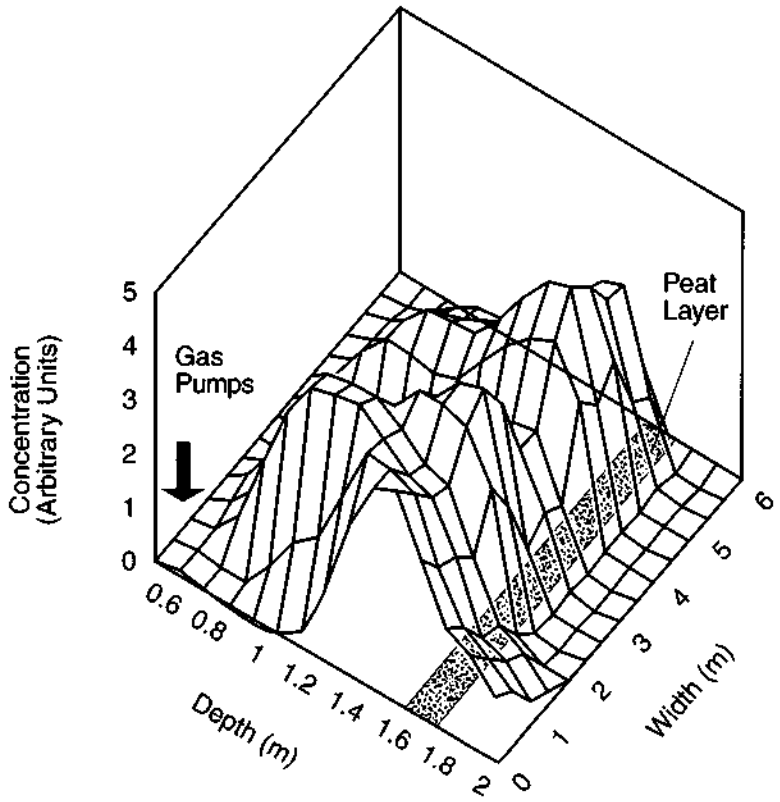


Figure 20-7 Profile of diesel fuel contamination in a vertical soil layer near a leaking gas pump. (From J. Bublitz *et al.*, *Appl. Opt.* **34**, 3223 (1995).)

4. Absorption Methods

Absorption methods provide a sensitive class of methods of detection for specific gases. The usual method of absorption spectroscopy involves transmitting the beam through the sample. One requires a two-ended system, with a detector on the opposite side of the sample from the laser source. This may be inconvenient for remote monitoring.

One possibility for relieving this difficulty is the use of a remote topographic reflector. One may use reflection from some feature present in the countryside, such as a large building or a hill, to direct light back to the position of the source. This can lead to a single-ended system in which the transmitter and receiver are located at the same position. This does, however, require that a suitable scattering feature be available.

Another variation of absorption spectroscopy for remote gas monitoring involves differential optical absorption spectroscopy (DOAS). DOAS relies on determining

the differences between local maxima and minima in the absorption spectrum of the gas being probed. DOAS seems to have been used mostly in the ultraviolet region of the spectrum. It requires a source that can cover a relatively large spectral range. DOAS has been demonstrated using laser sources, but it probably has been used more with broadband sources, like high-pressure xenon lamps.

With any of the foregoing absorption techniques for remote sensing, one obtains measurements of concentration integrated over the length of the absorbing path. The use of differential absorption laser (DIAL) backscattering techniques removes this difficulty and provides range resolution in a single-ended system. It also preserves the high sensitivity of absorption measurements and the specificity for particular gases. DIAL spectroscopy uses ambient atmospheric particulates and molecules as a distributed reflector to provide a return signal. The laser is tuned alternately on and off a characteristic molecular absorption peak to provide sensitivity for a specific gas. The gas concentration is determined from the relative strength of the return signal for on-peak and off-peak operation.

The essential feature of the DIAL technique is that it uses Rayleigh and Mie scattering in the atmosphere to provide reflection of light back to the position of the source, so that no physical reflector is needed. The tuning of the source wavelength on and off the peak of the absorption line provides high sensitivity and makes the method self-calibrating. If a short laser pulse is used, monitoring the return as a function of time after the pulse provides range resolution.

DIAL spectroscopy appears to be the leading candidate for laser-based remote monitoring of the environment. There have been many demonstrations of its use for remote sensing of a wide variety of materials. Some commercial DIAL systems are becoming available.

Figure 20-8 illustrates mapping of the distribution of SO_2 in a horizontal plane over the area of a zinc works, as determined using DIAL spectroscopy. The laser was located at the coordinate origin, off the right-hand side of the diagram. The measurements were performed using a Nd:YAG pumped dye laser as the source. The laser was tuned between the wavelengths of 296.17 and 297.35 nm to provide on-peak and off-peak conditions. The plume of SO_2, traveling with the wind, is mapped with a claimed uncertainty around 2 percent.

As another example, NASA has used an airborne ultraviolet DIAL system to monitor ozone concentrations [2]. The system uses two dye lasers pumped by Nd:YAG lasers, to generate 301 nm light for the on-peak probe and 311 nm light for the off-peak probe. The system has been flown on many missions to map the concentrations of ozone in the troposphere and stratosphere around the world, particularly the depletion of ozone in the arctic regions.

Mobile DIAL spectroscopy systems have been developed, suitable for remote monitoring of pollution at changing sites. DIAL spectroscopy has become an established technique for atmospheric monitoring. It is used for applications like industrial site inspection, monitoring of waste cleanup, enforcement of air quality regulations, and environmental research.

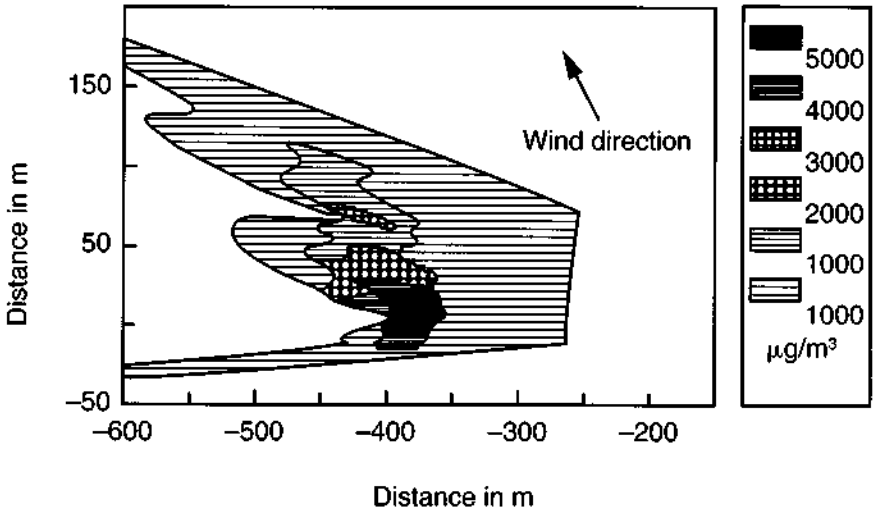


Figure 20-8 Distribution of SO_2 in a horizontal plane over the area of a zinc works, as determined with DIAL spectroscopy. (From U.-B. Goers, Laser remote sensing of sulfur dioxide and ozone with the mobile differential absorption lidar ARGOS, *Opt. Eng.* **34**, 3097 (1995).)

5. Summary

There have been many demonstrations of the application of lasers to remote monitoring of atmospheric parameters. Many different methods have been employed. Table 20-2 compares the features of the different methods we have described.

Some commercial systems for laser monitoring of the environment have become available. Models that are being offered make use of a variety of laser types, including Nd:YAG (both 1064 and 532 nm devices), excimer, tunable dye, tunable semiconductor, and CO_2 lasers. Systems have been designed for a number of specific applications, including measurement of smokestack opacity, monitoring of ozone concentration, leak detection, and cloud mapping. Although commercial laser-based environmental monitoring equipment is still in a relatively early stage, it appears to be developing steadily.

Remote sensing of atmospheric parameters involves directing a laser beam into the atmosphere. Issues of laser eye safety become important. The requirements for eye safety may limit the laser power that can be used in a probing beam. The relatively low values for maximum permissible eye exposure in the visible and ultraviolet may restrict the use of high-power pulsed lasers in these spectral regions. The optimum wavelength from the point of eye safety would be in the infrared, at wavelengths greater than 1.5 μm. The widespread application of laser technology for remote atmospheric monitoring will have to be compatible with public safety.

Although many workers have been active in this field for a number of years and have shown the capability for obtaining remote measurements of air pollutants and

Table 20-2 **Comparison of Laser-Based Remote Monitoring Methods**

Method	Identification of gases	Range resolution	Ease	Sensitivity
Optical radar	No	Yes	Single-ended	Moderate
Raman scattering	Yes	Yes	Single-ended	Low
Resonance fluorescence	Yes	Yes	Single-ended	Moderate
Absorption				
Transmission	Yes	No	Double-ended	High
Topographic	Yes	No	Single-ended, but needs reflecting feature	High
DOAS	Yes	No	Double-ended	High
DIAL	Yes	Yes	Single-ended	High

other atmospheric constituents, this technology is still largely in development. It has been useful as a research tool. Global scale measurements by NASA have provided useful information about atmospheric composition. Concentrations of specific gases around specific industrial sites have been derived. The amount of work being performed in this area indicates that laser-based remote sensing of atmospheric quantities will continue to grow.

References

[1] G. E. Busch *et al.*, *Proc. Nat. Acad. Sci.* **69**, 2802 (1972).
[2] E. V. Browell, *Proc. IEEE* **77**, 419 (1989).

Selected Additional References

R. M. Measures, ed., *Laser Remote Chemical Analysis,* Wiley, New York, 1988.
E. R. Menzel, *Laser Spectroscopy: Techniques and Applications,* Marcel Dekker, New York, 1994.
Optical Sensing for Environmental Monitoring, Proceedings of an International Specialty Conference, Atlanta, GA, October 11–14, 1993, Air and Waste Management Assn., Pittsburgh, PA, 1994.
L. J. Radziemski, R. W. Solarz, and J. A. Paisner, eds., *Laser Spectroscopy and Its Applications,* Marcel Dekker, New York, 1986.

Chapter 21 | Chemical Applications

The development of intense tunable laser sources has opened up many new areas of research in photochemistry. As a scientific tool, lasers have been used in chemistry for spectroscopy and flash photolysis, in which the dynamics of chemical reactions are studied. Such studies have yielded much useful information about the progress of chemical reactions.

Perhaps the most significant use of lasers in chemistry has involved spectroscopic applications. High-resolution spectroscopy with a tunable laser has become a valuable analytical tool. Picosecond spectroscopy has allowed monitoring of the dynamics of chemical processes with extremely high temporal resolution. The use of lasers in chemical spectroscopic applications was described in the last chapter.

Lasers can be used, in principle, to drive chemical reactions in a chosen direction. The tunable laser is tuned to a wavelength corresponding to a selected molecular excitation. This can stimulate the chemical reaction to proceed in a direction different from the one it would take without the photoexcitation. This potential application is still in a very early stage of investigation.

Another significant possibility for laser photochemistry involves isotope separation. This process utilizes the slight shifts that occur in the molecular absorption spectrum for molecules containing different isotopes of a given element. The laser is tuned to the resonant absorption of one of the isotopes. Molecules containing that isotope participate in some reaction and may be separated. This process can have practical significance, particularly for the separation of uranium isotopes.

Early experiments in laser photochemistry were performed with the fixed-wavelength lasers available in the 1960s. Such experiments studied reactions in which there was a fortuitous coincidence between the laser frequency and a resonant molecular absorption frequency. Developments in tunable lasers that allow laser operation at any frequency within a broad range now permit the excitation and study of almost any desired molecular state.

The applications described in this chapter emphasize the use of tunable lasers. Tunable lasers were described in Chapter 4. Many chemical applications

have been investigated using tunable dye lasers, which for many years were the only easily available broadly tunable lasers. In recent years, development of tunable solid state lasers, like Ti:sapphire, has added new tunable sources suitable for chemical applications. The use of nonlinear optical effects, like frequency doubling and tripling and Raman shifting, and the use of optical parametric oscillators (OPOs) extend the range of wavelengths available for photochemical studies.

Photochemical applications commonly have employed tunable dye lasers operating in the ultraviolet, visible, and near infrared portions of the spectrum, Ti:sapphire lasers in the visible, and near infrared, OPOs in the visible and near infrared, and line-tunable CO_2 lasers, which can be tuned to any of a large number of closely spaced lines in the infrared near 10 μm.

The field of laser photochemistry is extremely wide and varied. In this chapter, we can discuss only a few examples to illustrate the capabilities of laser photochemistry. We will describe the use of lasers to initiate chemical reactions, to alter the course of chemical reactions, to monitor the dynamics of chemical reactions, and to separate isotopes of various elements.

A. Laser-Initiated Reactions

Chemical reactions often require an activation energy to be supplied. Endothermic chemical reactions may not occur without activation, because the reactant molecules must be very close together to react. At small separations, repulsive forces between molecules come into play, and energy is needed to overcome these forces. Thermal energy has traditionally been used. The rate of chemical reaction increases with heating. But the internal energy associated with molecular vibrations can also be useful for activation. In fact, vibrational molecular energy appears even more conducive to promoting chemical reactions than thermal translational energy.

Thus, a relatively new approach to chemical reactions involves selective excitation of the vibrational states in reactant molecules by absorption of laser light tuned to the resonant absorption of specific states. This approach appears to offer advantages compared with exciting the molecule as a whole by providing thermal energy, most of which occurs as translational energy. The selective excitation of vibrational states may be more efficient in producing a reaction than the same amount of energy supplied as heat.

We will consider two examples. In one case [1], reactions were initiated with mixtures of the gases boron trichloride and acetylene. There is an overlap between the emission spectrum of CO_2 lasers and the absorption of boron trichloride. The absorption in the region near 10.6 μm excites an asymmetric stretching vibration of the boron trichloride molecule. The absorption of laser radiation and subsequent excitation of this vibrational level causes interaction between boron trichloride and acetylene. Acetylene, which is transparent to the CO_2 laser light, usually does

not react with boron trichloride. But under irradiation by pulsed CO_2 laser light, the following reaction occurred:

$$BCl_3 + C_2H_2 \rightarrow HCl + HC_2BCl_2 \tag{21.1}$$

In this laser-induced photochemical reaction, the dominant role belonged to those vibrational levels of molecules excited by absorption of laser light.

In a second example [2], CO_2 laser radiation at a wavelength of 10.6 μm interacted with the system N_2F_4–NO. Under ordinary circumstances, the N_2F_4 does not react with NO. But under irradiation by the CO_2 laser, the vibrational states of the N_2F_4 molecule were excited. Chemical reactions occurred to produce NF_3, NFO, N_2, and F_2. The percentage of each product depended on the experimental conditions, including gas pressure and laser irradiance.

Although many examples of laser-initiated reactions have been demonstrated, the field is still immature and does not appear to have become commercially viable as yet. One factor is the relatively high cost of providing the laser photons. The best candidates for commercial exploitation may be very efficient catalytic reactions.

Purification of materials is another possible application of laser chemistry. One potentially important purification process that has been demonstrated involves silane, SiH_4. This application is significant because silane is used as the starting material for making the high-purity silicon used in microelectronics and in solar cells.

Light from an ArF excimer laser operating at a wavelength of 193 nm interacted with impurities such as PH_3, AsH_3, and B_2H_6 in gaseous silane [3,4]. The reaction apparently involved photolysis of the impurity molecules, with resultant breakup of the molecules. The impurities were selectively removed from the silane. The remaining silane was of high purity. It was possible to obtain impurity levels below one part per million. These gaseous impurities are difficult to remove by other means. Removal of the impurities is an important step in improving the purity of silicon wafers that are epitaxially grown from silane.

Substitution of lasers for conventional light sources used to stimulate photochemical applications represents another type of laser-driven chemical reaction. An example involves the synthesis of vitamin D. The usual photochemical process begins with an isomerization reaction that converts a starting compound, 7-dehydrocholesterol, to previtamin D_3. The previtamin D_3 is later converted to vitamin D. The photochemical process for production of the previtamin is conventionally driven by a mercury arc source and has low yield. Experimental studies [5] have shown that a laser-induced process suppressed competing side reactions and increased the yield to 90 percent. The laser process used two-step irradiation with KrF laser light at 248 nm and nitrogen laser light at 337 nm. This substitution of light sources provided a substantial improvement for the synthesis of vitamin D.

B. Laser-Altered Reactions

Lasers provide the possibility of selective excitation of chosen excited states of atoms and molecules, or selective breaking of specified chemical bonds. This capa-

bility can be attractive for stimulation of desired chemical reactions. It may be possible to tune a laser to a molecular resonance so that only certain chosen chemical bonds are excited.

Specific vibrational states of molecules may be excited by absorption of laser light tuned to the proper frequency. The application of lasers to produce vibrationally excited molecules offers the possibility of selective stimulation of reactions involving only the excited chemical bond. An excited molecule may undergo chemical reactions along different pathways than if it were not excited, or one could drive chemical reactions selectively by excitation of reaction channels that are not normally available. A long-standing goal for laser chemistry has been to influence the course of chemical reactions so as to break selected chemical bonds and thus to change the relative yields of products of the reactions.

Although extensive research has been carried out, there are few good examples of chemical reactions that have been driven selectively with lasers [6], and apparently none have reached commercial status. Many of the studies have emphasized the excitation of vibrational energy levels corresponding to one or more quanta of vibrational energy. Selective excitation of vibrational states in a specified bond could drive the course of the reaction in a direction different from that produced by thermal energy.

But there have been difficulties in developing such bond-selective chemistry. Perhaps the greatest difficulty is that the energy remains localized in a specific vibrational mode for a very short time, and then it is transferred rapidly to other modes. A second difficulty is that the efficiency with which the vibrational energy of a molecule influences the course of its reactions is limited.

One recent good example of bond-specific chemistry driven by absorption of laser light has involved the gas phase reactions [7,8]

$$H + HOD \rightarrow OD + H_2 \tag{21.2}$$

$$H + HOD \rightarrow OH + HD \tag{21.3}$$

In these reactions, hydrogen atoms produced in a microwave discharge react with water molecules having two different hydrogen isotopes. The reaction has two possible pathways, as indicated by the two equations. Pulses of light from a frequency-doubled dye laser were used to monitor the OH and OD fragments via laser-induced fluorescence. Laser light at 724 nm wavelength from a tunable Ti:sapphire laser was absorbed by the HOD molecule, raising the molecule into a vibrationally excited state with four quanta of O—H stretching excitation. Then, reaction (21.2) occurred about 200 times as frequently as reaction (21.3). But when the wavelength was changed to 781 nm, exciting the molecule with five quanta of O—D stretching vibration, reaction (21.3) occurred about 220 times as frequently as (21.2). In both cases, the bond that was excited by the absorption of laser light reacted much more strongly. Thus, vibrational excitation could be used to control the reaction to produce either product almost exclusively. Without vibrational excitation, the two reactions occur at about the same rate. Apparently, in this set of reactions, the vibrational

energy remains localized in the specific bonds, at least for a long enough time to promote preferential reaction of that bond.

In another experiment involving the same set of reactions [9], excitation of a single quantum of O—H stretching vibration by an OPO operating near 2.63 μm wavelength drove reaction (21.2). But excitation of a single quantum of O—D vibration, with the OPO operating near 3.57 μm, drove exclusively reaction (21.3). These results indicate that selective excitation of a specified bond in a polyatomic molecule, even by a single quantum of vibrational energy, can lead to preferential reaction involving that bond.

An example of a bond-specific reaction that involves two chemically distinct channels, rather than isotopically distinct channels, is the photodissociation of isocyanic acid (HNCO) [10]. This molecule can be dissociated by breaking either the C—N bond or the H—C bond, according to the reactions

$$\mathrm{HNCO} \rightarrow \mathrm{NH} + \mathrm{CO} \tag{21.4}$$

$$\mathrm{HNCO} \rightarrow \mathrm{H} + \mathrm{NCO} \tag{21.5}$$

The demonstration of the selective control of the reaction pathway by vibrational excitation of the molecule involved use of four lasers: a Nd:YAG-pumped dye laser operating at 759 nm to excite the vibrations of the HNCO molecule, a frequency-doubled Nd:YAG-pumped dye laser operating at 340 nm to dissociate the molecule, and two additional tunable ultraviolet lasers to excite laser-induced fluorescence of the NCO and NH fragments for monitoring purposes.

When the HNCO molecule absorbed the 759 nm light, it was excited with four quanta of N—H stretching vibration. Then, the photodissociation occurred preferentially with breaking of the N—H bond, and reaction (21.5) was enhanced relative to reaction (21.4) by a factor at least 20. These results demonstrate the ability to cleave a selected chemical bond by means of laser-induced vibrational excitation.

These demonstrations of mode-selective chemistry have illustrated the principle of control of chemical reactions, but they have not yet led to practical chemical process control. The most important application has been as an experimental tool to elucidate the dynamics of chemical reactions. There remain important unanswered questions about whether the control of reactions may be extended to larger molecular systems and whether the same type of control may be established for reactions in condensed media, in which most chemical processing occurs. The exploitation of bond-specific laser-controlled chemistry to drive commercial processes and to enhance yields of specific products remains in the future.

Another important area of investigation involving laser-altered photochemistry is dissociation of molecular species by absorption of one or several photons [11,12]. In particular, photodissociation by lasers has been of interest because of the possibility of breaking a specific bond in a molecule. This again leads to the possibility of driving the dissociation reaction in selected directions, to produce breakup in specific controllable reaction channels.

The dissociation energy of many molecules is around 3 eV. If one attempts to dissociate the molecule by absorption of a single photon, the wavelength will lie in the

ultraviolet. Because the photodissociation occurs very rapidly, on a subpicosecond time scale, the molecule is relatively insensitive to the exact photon energy, and more than one excited state may be produced. The dissociation can proceed by more than one pathway, and the desired goal of control of the dissociation pathway is not achieved.

Alternatively, one may supply the energy through absorption of a larger number of lower-energy photons. If one uses an infrared laser with a photon energy around 0.1 eV, about 30 photons would have to be absorbed to produce dissociation. Figure 21-1 shows the molecular binding energy for a polyatomic molecule as a function of interatomic distance. The horizontal lines represent bound excited states of the molecule, that is, the vibrational states of the molecule. If the molecule absorbs a number of photons rapidly, it will reach an energy above the dissociation limit, and the atoms can move apart to indefinitely large distances as the molecule dissociates.

The lowest energy levels are approximately evenly spaced. If a high-power laser is tuned to the wavelength corresponding to this spacing, a number of photons will be absorbed in rapid succession and the molecule will make many transitions, rising up the ladder of excited states, as indicated by the vertical arrows. For the first steps up the ladder, the laser wavelength must be tuned to the proper value to drive the molecule up the evenly spaced steps. Above the first steps, the spacing becomes unequal, because of anharmonicity of the potential well. But the presence of additional rotational and translational states tends to fill the gaps between the vibrational states, forming a quasi-continuum. Thus, a tuned infrared laser can still drive the system upward to dissociation, as indicated by the arrows. Absorption can continue even

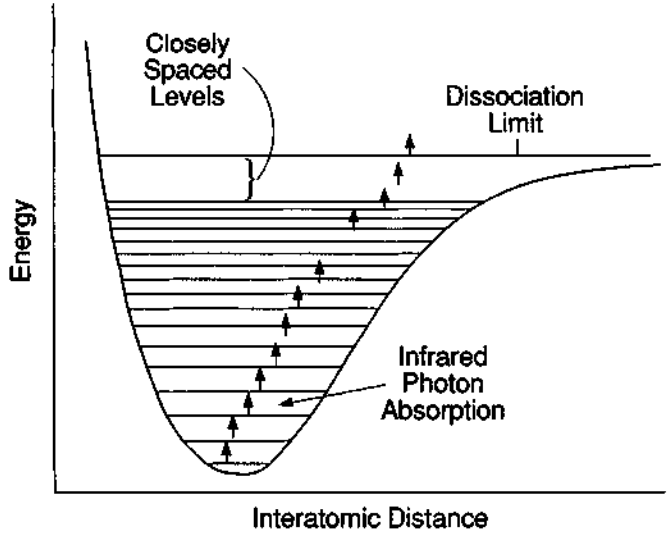


Figure 21-1 Vibrational energy levels for a polyatomic molecule as a function of interatomic distance. The vertical arrows indicate transitions induced by absorption of laser photons.

though the laser wavelength no longer exactly matches the spacing of the vibrational states.

One example of unimolecular dissociation of this type involved breakup of gaseous SF_6 according to the reaction [13]

$$SF_6 + n\,h\nu \rightarrow SF_5 + F \tag{21.6}$$

where the integer n represents the number of laser photons needed to supply the dissociation energy. In this case, 35 photons were required.

This reaction was driven by a CO_2 laser line-tuned to the wavelength of 10.61 μm, corresponding to the spacing of the lowest few states of the SF_6 ladder. The laser was a high-power TEA laser with pulse duration around 100 nsec, so that there was no time for energy transfer by collisions.

This example shows the potential for breakup of individual molecules by a tuned laser. Many such unimolecular dissociation reactions have been studied and have provided a wealth of data on molecular structure and reaction dynamics. Although there have been some examples indicating bond-selected reactions, there has apparently as yet been no commercial exploitation of them.

C. Laser Monitoring of Chemical Dynamics

Laser technology has made significant contributions in probing the dynamics of chemical reactions, especially high-speed reactions. We shall present a few examples, out of many possible, that demonstrate the versatility of laser-based techniques for investigating chemical reactions.

1. PICOSECOND PHOTOCHEMISTRY

Laser diagnostic procedures have been especially effective in probing the dynamics of chemical reactions on a very rapid time scale, of the order of picoseconds, and yielding information that would not otherwise be available. Figure 21-2 shows one configuration that has been employed. The laser pulse is split into a pump pulse and a probe pulse. The pump pulse irradiates the sample and initiates the photochemical reaction. The probe pulse is split into a number of beams, which are delayed by different amounts by a series of glass plates. The sample is thus probed by a series of very short pulses. The absorption of specific reaction products may be monitored as a function of time. The temporal evolution of the reaction products is thus revealed on a very rapid time scale.

Many picosecond studies of chemical reaction mechanisms have been performed. An example of picosecond probing of the formation and decay of the biophysical compound prelumirhodopsin has been presented in Chapter 20. Picosecond techniques have been useful for many chemical applications, such as studying the excited states of biological systems and determining the vibrational and electronic relaxation processes of large molecules.

Most such work has been done using dye lasers as the picosecond source, but now studies utilizing tunable Ti:sapphire lasers are becoming common.

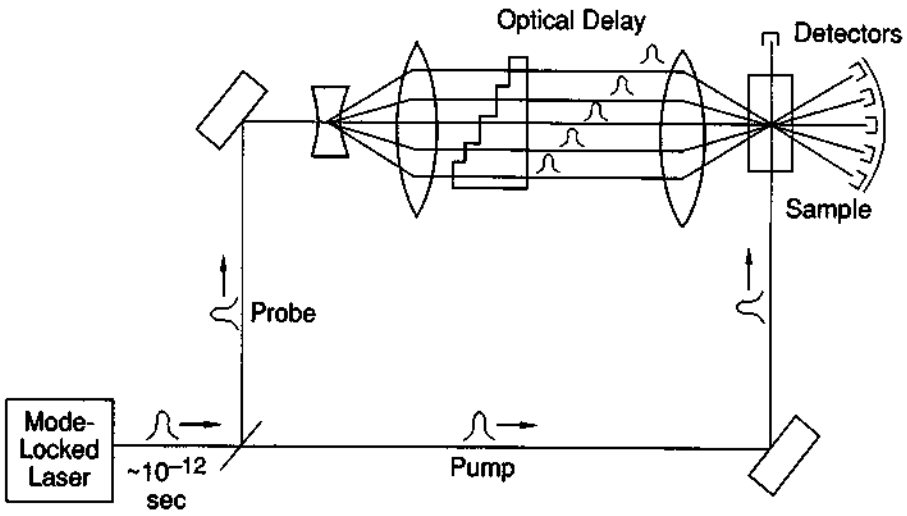


Figure 21-2 Apparatus for picosecond monitoring of chemical reaction products.

2. *FLASH PHOTOLYSIS*

Flash photolysis also provides a technique for the study of rapid chemical reactions. It involves the breakup of a molecular species by an intense flash of light. Then, the absorption lines of the resulting products are monitored to determine the identities and concentrations of the reaction products as a function of the time after the flash. In this way, the course of the chemical reactions may be monitored.

Flash photolysis has two basic variations. In the first technique, a second flash with a variable time delay is used to provide a spectroscopic continuum against which to record the absorption spectrum of the reaction products produced by the first flash. A photographic record of the spectrum over a broad spectral range is obtained. This technique is usually used for exploratory work, when one does not know the nature of the chemical species that will appear. Once the chemical species are identified and their absorption lines are known, the second technique may be employed. The second flash is replaced by an intense continuous background source. Observations are made at a single wavelength using a monochromator and detector. Alternatively, a tunable continuous laser may be used. The second technique provides better detailed kinetic studies, because it gives a complete time history of the variation of the absorption due to a particular species.

Flash photolysis has been used to generate photochemically and to study spectroscopically a wide variety of free radicals and excited molecules. The minimum temporal resolution in flash photolysis is limited mainly by the duration of the initiating flash. Before lasers were available, this minimum resolution was in the microsecond regime. The availability of lasers with short pulses has reduced this time scale substantially. The earliest laser experiments used Q-switched ruby lasers with pulse durations in the 20 nsec regime [14,15]. This had the immediate effects of extending

the capabilities of flash photolysis from the microsecond to the nanosecond time scale, and of providing a new very intense source for the initiating pulse.

Now the availability of picosecond and femtosecond pulsed tunable lasers has extended the capabilities of flash photolysis much further. Flash photolysis work is routinely carried out with picosecond resolution. Flash photolysis has been employed to analyze the structure of excited states of molecules, and to determine the kinetics of large organic molecules in solution. Kinetic monitoring has yielded data such as fluorescent lifetimes, oxygen quenching of singlet and triplet states of molecules, and absolute values for absorption coefficients. The availability of laser-based flash photolysis has been a valuable tool in analytic spectrochemistry.

3. *STATE-TO-STATE CHEMISTRY*

The availability of tunable lasers makes it possible to produce molecules in selected internal quantum states, involving electronic, vibrational, or rotational energy. After the molecules in selected states have undergone a chemical reaction, one may use tunable lasers to determine the internal quantum states of the product molecules, through absorption or fluorescence. The result is a capability to perform "state-to-state" chemical analysis, determining the cross sections and reaction rates for producing product molecules in specified quantum states from starting molecules in selected quantum states. This can provide valuable information to chemists about the pathways that a chemical reaction takes.

Studies of state-to-state reactions have been performed on a wide variety of chemical systems. We discuss one specific example, the photodissociation of water, to illustrate typical techniques. In this work [16], a dye laser pumped by a Nd:YAG laser emitted pulses at a wavelength near 700 nm. This laser irradiated a low-pressure sample of water vapor and produced excitation in the water vapor. A second pulse in the ultraviolet region, produced by a second laser system, irradiated the sample 30 nsec after the exciting pulse, and this caused photolysis. The second laser used the fourth harmonic of Nd:YAG at 266 nm or light at 239.5 or 218.5 nm produced by Raman shifting or frequency-mixing effects.

The fragments from the dissociation reaction were probed by light from a third laser, a frequency-doubled nitrogen-laser-pumped dye laser operating at 308 nm. This device produced laser-induced fluorescence, which could be analyzed to obtain the rotational and vibrational states of the products.

The results yielded the relative populations of the rotational states of the OH fragment resulting from the photolysis, for a number of different rotational states of the initial water molecule. The distribution was found to depend strongly on the initial state of the water molecule. Similarly, the population of vibrationally excited states in the OH fragment varied as the vibrational state of the parent water molecule was changed. The results yielded a full resolution of the states involved in the photolytic reaction and allowed determination of the role of initial rotational and vibrational excitation in the dynamics of the reaction. This information could not have been obtained without the use of laser technology.

D. Isotope Separation

1. PRINCIPLES

Isotope separation may be one of the most important of the photochemical applications of lasers. It has particular significance because of the possibility of the separation of uranium isotopes. And procedures for the separation of the isotopes of many other elements have also been developed. Such processes are important for the production of isotopically enriched elements for medical, industrial, and scientific applications.

Laser photochemical isotope separation involves the use of a laser tuned to a resonant absorption of a molecule containing an isotope of the desired atom. The spectral absorption lines of atoms and molecules have slightly different wavelengths, depending on their isotopic constitution. The difference between the wavelengths for atoms or molecules with different isotopes is called the isotope shift. Because of the isotope shift, light energy may be deposited by absorption in an atom or molecule with one isotopic composition, but not in the same atom or molecule with a different isotopic composition. There are isotopic shifts in the absorption spectra of many atoms and molecules, but the values of the isotope shift are generally small. Because of the smallness of the shifts, very narrowband light sources are required to excite the absorption of one specific isotope. Only the laser can produce both the required narrow linewidth and the required high power for practical photochemical separation of isotopes. The existence of isotope shifts and the availability of tunable lasers with narrow linewidths leads to the possibility of exciting and separating designated isotopes with laser radiation.

Absorption of laser light by molecules containing the desired isotope leads to a chemical or physical reaction involving only the molecules containing that isotope. The products of the reaction may be separated to yield a product that is enriched in the desired isotope.

There are several things needed for laser-assisted isotope separation. These include

- A shift in the absorption spectrum between different isotopes of the desired element
- A tunable laser that can be tuned to excite only one of the isotopes
- A chemical or physical process that acts only on the excited species and separates it from the unexcited molecules
- The absence of any scrambling process, such as thermal excitation, that can undo the effect of exciting only one of the isotopes

The material may be either atoms of the desired element or molecules containing the desired isotopic species. For atoms, the spectral lines of interest are in the electronic structure and tend to lie in the visible or ultraviolet regions. For molecules, the spectra of interest are vibrational or rotational in origin, and the lines of interest tend to lie in the infrared. Some characteristic isotope shifts for some absorption lines for both atoms and molecules are presented in Table 21-1.

Table 21-1 Isotope Shifts

Atoms	Wavelength (nm)	Isotope shift (cm^{-1})	Doppler linewidth* (cm^{-1})
^{6}Li–^{7}Li	323.26	0.35	0.026
^{10}B–^{11}B	249.77	0.175	0.44
^{200}Hg–^{202}Hg	253.65	0.179	0.04
^{235}U–^{238}U	424.63	0.280	0.055
Molecules	**Wavelength (μm)**	**Isotope shift (cm^{-1})**	**Rotational bandwidth (cm^{-1})**
$^{10}BCl_3$–$^{11}BCl_3$	10.15	9	25
$^{14}NH_3$–$^{15}NH_3$	10.53	24	71
$^{48}TiCl_4$–$^{50}TiCl_4$	20.06	7.6	18
$^{235}UF_6$–$^{238}UF_6$	16.05	0.55	16

*At a temperature high enough to produce a vapor pressure of 1 Torr.

The isotope shift must be compared with the linewidth of the spectral line. If the lines of the two isotopes overlap, a selective excitation of only one isotope will be impossible. Because spectral lines in solids and liquids are generally broad, usually only gaseous media are considered for photoseparation of isotopes. The linewidth for atomic lines in gaseous media at low pressure is the Doppler linewidth, proportional to the square root of the temperature. The isotope shifts of some of the atomic species in Table 21-1 are larger than the Doppler width, which is shown for a temperature high enough to produce a vapor pressure of 1 Torr for the element. In general, the isotope shift is larger than the Doppler width only for very light or for very heavy atoms.

For molecules, the relevant bandwidth is the width of the rotational substructure. This is also given in the table. This width is usually larger than the isotope shift. This often makes isotope separation using molecular species difficult, although some methods have been suggested for evading the problem. One such method is to use absorption lines arising from simultaneous excitation of more than one vibrational quantum. Flow cooling of gaseous species can also be used to narrow the linewidth.

There are several possible forms that the separation process can take. Some of these are illustrated in Figure 21-3.

Two-step photoionization The atom of the desired isotope is excited by absorption at one wavelength. Then excited atoms are ionized by absorption of light at a second wavelength. Unexcited atoms do not absorb the second wavelength. This process is shown schematically in Figure 21-3a. The ions may be extracted electrically. This process has the desirable feature of easy and efficient extraction. However, it is not conservative of light energy, because the second wavelength is usually short and the associated photon energy is high.

Two-step photodissociation A molecule containing the desired isotope is excited by absorption of one wavelength, whereas molecules with other isotopes

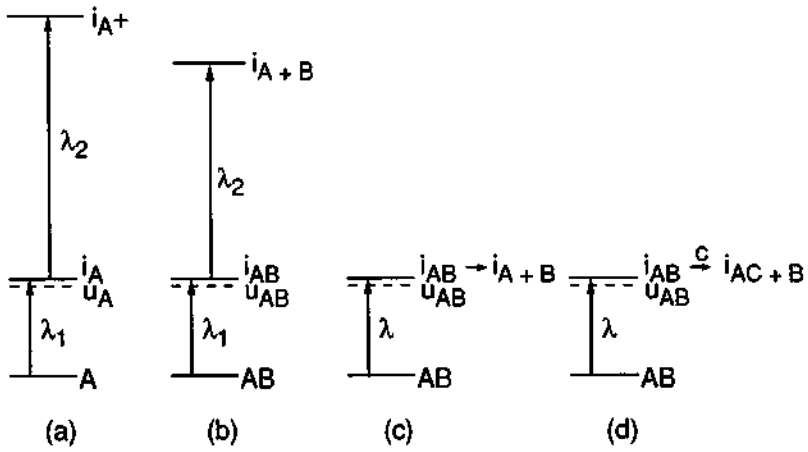


Figure 21-3 Four of the possible isotope separation processes for a molecule, designated AB. In this figure, the left superscript i denotes an isotopically selected species and the superscript u an unselected species. The laser wavelengths are λ_1 and λ_2. (a) Two-step photoionization. (b) Two-step photodissociation. (c) Photopredissociation. (d) Chemical reaction of excited species.

are not. Then, the excited molecules are broken up by absorption of a second wavelength. This process is sketched in Figure 21-3b. The dissociation products must be extracted by some chemical or physical means. The dissociation products may be free atoms of the selected isotope or possibly fragments of the original molecule containing the desired isotope. This process, like two-step ionization, is energetically expensive.

Photopredissociation Photopredissociation occurs in some molecules that are excited into metastable states. The excited molecule has energy levels corresponding to unbound states, which cross the bound excited states. An excited molecule can make a crossover into the unbound states and dissociate. This process is shown in Figure 21-3c. This approach is more conserving of energy, because no second high-energy photon is needed, but it depends on the availability of molecules with suitable energy level structure.

Chemical reactions of excited molecules In this approach, the excited molecules undergo a chemical reaction, whereas unexcited molecules do not. Such selective laser-stimulated reactions have already been discussed earlier in this chapter. The process is shown schematically in Figure 21-3d. The reaction product must be separated after the reaction. This process is also conserving of energy. But at present, identification of suitable laser-excited reactions is limited. Wide use of this method will depend of identification of suitable reactions for specific molecules.

Deflection of an atomic beam This method relies on resonant absorption and reemission of laser light by the desired isotopic species in an atomic beam.

The absorption causes the atoms to gain momentum in the direction perpendicular to the beam. Two spatially separated atomic beams result, one containing the desired species and the other containing the unselected isotopes. The atoms are separated in space, so that collection is simple. The method should be very conserving of energy, because the reemitted light could be collected and reused. In principle, the method should be applicable to any heavy element. It has been demonstrated with barium.

Multiple-step infrared photodissociation In an idealized picture, the vibrational energy levels of a polyatomic molecule will be spaced equally in energy. If a laser is tuned to the proper frequency, absorption can take place sequentially, one step at a time, driving the molecule up the ladder of equally spaced states until it has absorbed enough energy to dissociate. This is shown schematically in Figure 21-4. Slight isotope shifts in the vibrational levels allow isotopic selection. In practice, the energy levels are not equally spaced, so that one would not expect a tuned laser to drive the molecule more than a few steps up the ladder. Yet it has been used successfully, for example, to separate isotopes of sulfur, by irradiation of sulfur hexafluoride by a CO_2 laser, using the same interaction already discussed in connection with Equation (21.6). The laser is tuned to the absorption of the selected isotope and drives molecules of that isotope up the first few steps of the ladder, whereas molecules with the unselected isotope are not excited. Above the first steps, anharmonicity of the potential well makes the spacing unequal. But the presence of additional rotational and translational states tends to fill the gaps between the vibrational states, so that molecules containing the selected isotope will continue to absorb energy and will eventually dissociate.

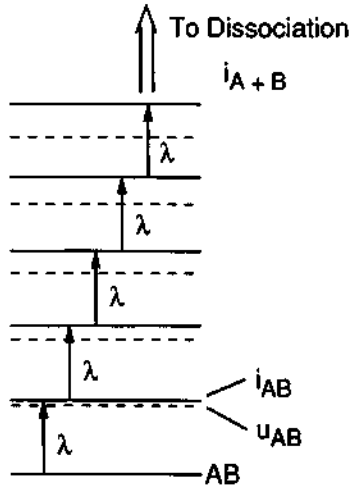


Figure 21-4 Multiple-step infrared dissociation for a molecule, designated AB. In this figure, the left superscript i denotes an isotopically selected species and the superscript u an unselected species. The laser wavelength is λ.

There have been many other suggested methods for isotope separation. However, the six methods listed here illustrate the principles of laser-driven isotope separation.

2. *URANIUM ISOTOPE SEPARATION*

The most significant application of laser-based isotope separation involves uranium, separating ^{235}U from ^{238}U. The separation is important because isotopically enriched ^{235}U is needed for light-water reactors, used for electrical power generation in the United States and elsewhere. Natural uranium contains 0.71 percent ^{235}U, which is the fissionable isotope necessary for reactor operation. The remainder in nonfissionable ^{238}U. Commercial light-water reactors require about 3 percent ^{235}U. Thus, natural uranium must be enriched before use as fuel in light-water reactors.

Conventional production of uranium enriched in ^{235}U employs gaseous diffusion or gas centrifuges. These processes rely on the mass differences between the uranium isotopes, with the lighter atoms moving faster than the heavier ones. Because the mass difference is small, the separation process requires many stages. Gas diffusion plants are very large, expensive to operate, and require very large amounts of power. Gas centrifuge plants require less power but are also very large and expensive. A successful laser-based separation process would increase the supply of ^{235}U available for use in nuclear reactors and reduce its cost substantially.

Although a number of approaches to uranium isotope separation have been studied, the leading contender is the Atomic Vapor Laser Isotope Separation (AVLIS) process, which has been under development for a number of years at Lawrence Livermore National Laboratory [17]. AVLIS is a multistep photoionization process. The United States Department of Energy selected AVLIS in 1985 as the technology to meet future needs for production of enriched uranium.

In the AVLIS process, metallic uranium is vaporized by an electron beam to produce a stream of uranium vapor. Figure 21-5 shows the apparatus schematically. The uranium atoms are irradiated by dye lasers emitting orange-red light at wavelengths that will selectively ionize ^{235}U. Pulses from three tunable lasers are superimposed on the stream. The wavelengths of the lasers are tuned so that ^{235}U absorbs the light, whereas the ^{238}U does not. The ^{235}U atoms are raised to an excited state by absorption of one photon, and then these excited atoms are ionized by absorption of two more photons at different wavelengths. The wavelengths of the lasers are in the 600 nm region. The resulting ^{235}U ions are removed by an electromagnetic field and collected at the product collector. The ^{238}U isotopes pass through the collector because they are not ionized.

The development of the dye laser system for AVLIS has also been the subject of intensive work [18]. The dye lasers are pumped by copper vapor lasers. Both the pump lasers and the dye lasers are master oscillator, power amplifier (MOPA) chains. The output of the dye lasers is tunable over a broad range from 550 to 650 nm, covering the relevant wavelengths of uranium absorption. The center frequency of the laser is stable to ±30 MHz, so that overlap of the laser wavelength and the absorption wavelength can be maintained.

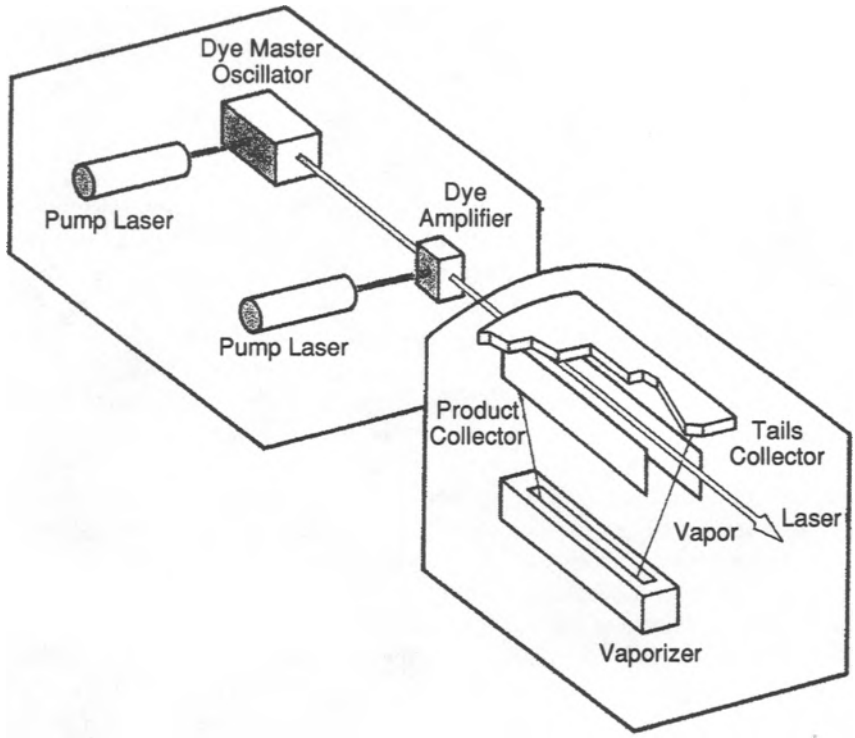


Figure 21-5 Apparatus for the AVLIS process. (From H.-L. Chen and J. I. Davis, *Photonics Spectra,* p. 59 (October 1982).)

The average power of a single dye laser chain exceeds 1300 W, in a near-diffraction-limited beam. The lasers are pulsed at repetition rates up to 10 kHz, with pulse durations in the 20–200 nsec regime. These lasers have been developed to provide reliable operation under conditions suitable for use in a uranium enrichment plant.

After more than two decades of development, AVLIS has reached the status of an industrial-scale process. In 1995, the AVLIS process was transferred from the Lawrence Livermore National Laboratory to an industrial organization for commercial development.

3. *SEPARATION OF NON-URANIUM ISOTOPES*

Separation of isotopes of elements other than uranium is also important for providing isotopically enriched materials for medicine, research, and industry. Isotopically enriched chemicals are widely used in medicine, for both diagnostic and therapeutic purposes. Radioisotopes are often used as tracers in industrial processes.

The cost of isotopically pure materials has been high. Laser techniques could reduce costs considerably. In most cases, only relatively small amounts of enriched material are required. But in some cases, large-scale separation is necessary. One

example is the separation of deuterium (^{2}H) from the much more abundant light hydrogen (^{1}H). Deuterium is used in the form of heavy water as a moderator to slow fission neutrons in reactors. A second example is the separation of the light isotope of boron (^{10}B) from the more abundant ^{11}B. The light boron isotope is used in reactor control rods because of its very large neutron cross section.

Laser isotope separation techniques have been demonstrated for many elements, including hydrogen, boron, carbon, nitrogen, oxygen, silicon, sulfur, chlorine, selenium, bromine, molybdenum, barium, osmium, mercury, and a number of the rare earth elements. In this section, we will consider a few representative examples to show the range of the laser-based isotope separation techniques.

The first example involves the use of multiphoton infrared photodissociation to increase the concentration of ^{34}S in SF_6. Mixtures of sulfur hexafluoride and hydrogen were irradiated with focused CO_2 laser radiation, delivering 1–2 J in a 200-nsec-duration pulse, with a peak irradiance around 6 GW/cm^2 [19]. The laser could be line-tuned to the P(20) line, which matches the initial step of $^{32}SF_6$ in its vibrational manifold. The relevant absorption spectra of $^{32}SF_6$ and $^{34}SF_6$ are shown in Figure 21-6. Irradiation with the P(20) line leads to excitation of $^{32}SF_6$ but not of $^{34}SF_6$.

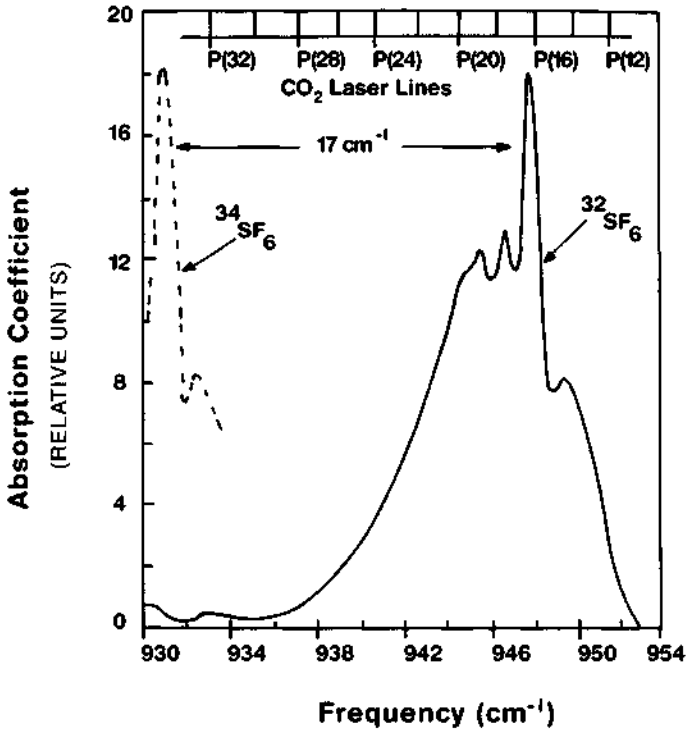


Figure 21-6 Absorption spectra of $^{34}SF_6$ and $^{32}SF_6$ in the infrared near 10.6 μm. The positions of several CO_2 laser lines are indicated. (From B. A. Snavely, at the IEEE/OSA Conference on Laser Engineering and Applications, Washington, DC, May 28–30, 1975.)

After irradiation, the residual sulfur hexafluoride was analyzed to determine its isotopic content. The original ratio $^{34}S/^{32}S$ was around 0.044. Figure 21-7 shows results for the composition of the starting material and for the residual SF_6 after 2000 and after 5000 pulses. These results indicate a substantial growth in the ratio of $^{34}S/^{32}S$ as a result of the irradiation. The ratio increases by a factor around 33 after 5000 pulses. The mechanism seems to be multiphoton dissociation of $^{32}SF_6$, with the liberated fluorine being scavenged by the hydrogen. The unselected $^{34}SF_6$ does not react and progressively becomes a larger component in the residual material.

A second example illustrates the use of photopredissociation to separate bromine isotopes. A mixture containing $^{79}Br_2$ and $^{81}Br_2$ was irradiated by a frequency-doubled Nd:YAG laser operating in the green near 558 nm [20]. The laser was tunable over a narrow range, about 1 cm^{-1}, and could be tuned to an absorption line of either $^{79}Br_2$ or $^{81}Br_2$. The resulting excited bromine molecules could dissociate by crossing over to an unbound state, which crosses the potential curve of the bound excited state. With the laser tuned to the absorption of $^{81}Br_2$, the reactions

$$^{81}Br_2 + h\nu \rightarrow {}^{81}Br_2^* \tag{21.7}$$

$$^{81}Br_2^* \rightarrow {}^{81}Br + {}^{81}Br \tag{21.8}$$

occurred. Chemical scavenging of the ^{81}Br then occurred via

$$^{81}Br + HI \rightarrow H^{81}Br + I \tag{21.9}$$

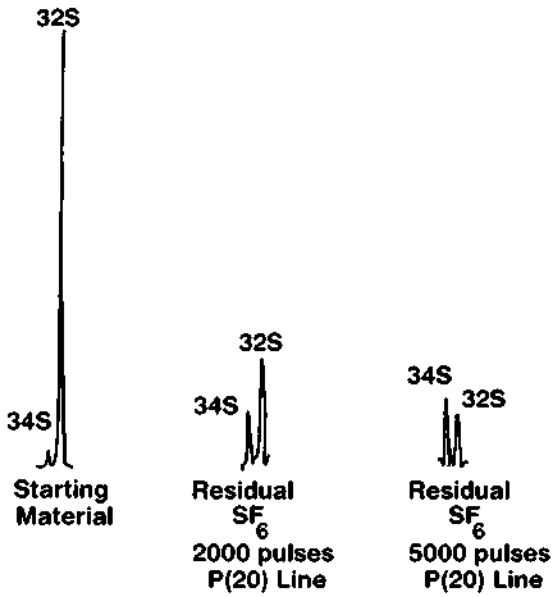


Figure 21-7 Laser enrichment of ^{34}S by CO_2 laser irradiation of SF_6. (From C. P. Robinson, *Ann. NY Acad. Sci.* **267**, 81 (1976).)

yielding a product that could be chemically separated. Enrichment from a natural abundance of 50 percent ^{81}Br to an abundance around 80 percent was demonstrated. The presence of possible scrambling reactions, such as

$$^{81}Br_2^* + {}^{79}Br_2 \rightarrow {}^{81}Br_2 + {}^{79}Br_2^* \tag{21.10}$$

limits the maximum abundance that can be achieved. The presence of this reaction yields $^{79}Br_2$ in the end product. Thus, the scavenging and collection processes must be completed quickly to avoid scrambling reactions.

A third example involved photodeflection of barium atoms [21]. In the photodeflection method, momentum is transferred from the laser beam to an atomic beam, traveling perpendicular to the laser beam. Each time an atom absorbs a photon, it acquires momentum $h\nu c$ in the direction of propagation of the photon, where $h\nu$ is the photon energy and c is the velocity of light. In the experiment, the atomic excitation was allowed to decay spontaneously. Because, on the average, the emission is isotropic, the momentum imparted to the atom by a large number of such emission processes is zero. Spatial separation of the desired isotope is achieved when the number N of absorbed photons is such that $Nh\nu c$ is larger than the transverse momentum spread of the atomic beam. One can regard the selected atoms as simply being pushed to the side, out of the atomic beam, by the momentum transfer from the light.

The experimental arrangement used barium metal, with the natural isotopic composition, evaporated in an oven. The vapor was collimated, and it intersected the laser beam at right angles. It was detected by a mass spectrometer. A 150-mW beam at a wavelength 553.57 nm from a tunable dye laser was used. The laser could be tuned to the correct wavelength for any of the barium isotopes, ^{134}Ba, ^{135}Ba, ^{136}Ba, ^{137}Ba, or ^{138}Ba. A separation of 10^{14} atoms per second was demonstrated, enriched by a factor about three over the natural abundance.

Another isotope separation procedure that has been extensively studied is the separation of tritium, ^3H, from both light water and heavy water. This separation is desirable both to remove radioactivity from the heavy water, D_2O, used as a moderator in heavy-water reactors, and to separate the tritium as fuel for thermonuclear fusion reactions.

Many groups have employed infrared laser photodissociation to separate tritium from both light water and heavy water. The usual procedure has been to use halogenated hydrocarbons (like CF_3H) as working molecules. These molecules can pick up tritium molecules by aqueous exchange with the original tritium-containing water. Then, these molecules are photodissociated by multiple photon absorption from an infrared laser beam. The wavelength is tuned to an absorption line of a molecule containing a tritium atom. Molecules containing hydrogen or deuterium atoms do not absorb. The tritium fragments are then scavenged by chemical or physical processes.

As an example, we consider the separation of tritium from deuterium [22]. A mixture of the compounds CCl_3T and CCl_3D in argon buffer gas was irradiated by the beam from an ammonia laser operating at a wavelength near 12.08 μm.

The study indicated enrichment of the tritium, originally present in very small percentages, by a factor greater than 15,000. The fraction of tritium in the original mixture was depleted, whereas the deuterium was not depleted. Other, similar studies using different working molecules have also demonstrated effective separation of tritium.

The AVLIS process, described previously with reference to uranium isotope separation, is also relevant for the separation of isotopes of many other elements. In one application, it was used to separate the isotopes of rubidium [23]. A collimated beam of rubidium atoms was irradiated by a tunable Ti:sapphire laser. The selected isotope was excited by absorption of light from the laser at a wavelength near 780 nm. The excited atoms were then photoionized by light from a frequency-tripled Nd:YAG laser at 355 nm. The photon energy at 355 nm was sufficient to ionize the excited atoms, but not atoms in the ground state.

As the frequency of the Ti:sapphire laser was tuned over a range around 10 GHz, the composition of the ion current changed. As the laser frequency was scanned successively over the absorption lines of ^{85}Rb and ^{87}Rb, the composition of the ion current, as monitored by laser-induced fluorescence, changed to correspond to the selected isotope. The fluorescence spectra demonstrated that the two rubidium isotopes were being selectively photoionized.

As the foregoing examples show, isotope separation processes usually do not yield complete separation of the one selected isotope; rather, they yield mixtures in which the amount of the selected isotope is increased over its natural abundance. Laser-based isotope separation processes provide sufficient enrichment for many practical purposes.

These processes are given as examples to show the operation of different aspects of a laser-based isotope separation process. There are a wide variety of photointeractions and chemical interactions that may be employed. The optimum separation process will vary from one element to another.

Although many examples of laser isotope separation have been demonstrated, this technology remains largely experimental. Use for industrial production of enriched isotopic species remains in the future, except for the case of uranium, for which an AVLIS process has been transferred to commercial development.

The examples of chemical processing and isotope separation procedures we have given represent only a small number of the many photochemical applications of lasers that have been studied. Laser photochemistry has been an extremely versatile tool for research and diagnosis. It has yielded substantial new information about reaction kinetics and the dynamics of chemical reactions. But laser processes have not yet made the transition to routine industrial chemical processing. There are few, if any, examples of laser processes with practical economic importance. The use of laser separation of uranium isotopes, which has been transferred from government laboratories to industry, may provide one of the first examples. The widespread use of laser technology for chemical synthesis and the selective control of chemical reactions has not yet been achieved, and these fields remain hopes for future development.

References

[1] N. V. Karlov, *Appl. Opt.* **13**, 301 (1974).
[2] N. G. Basov *et al.*, in *Fundamental and Applied Laser Physics* (M. S. Feld, A. Javan, and N. A. Kurnit, eds.), Wiley, New York, 1973.
[3] J. H. Clark and R. G. Anderson, *Appl. Phys. Lett.* **32**, 46 (1978).
[4] A. Hartford *et al.*, *J. Appl. Phys.* **51**, 4471 (1980).
[5] V. Malatesta, C. Willis, and P. A. Hackett, *J. Am. Chem. Soc.* **103**, 6781 (1981).
[6] H. R. Bachmann *et al.*, *Chem. Phys. Lett.* **29**, 627 (1974).
[7] A. Sinha, M. C. Hsiao, and F. F. Crim, *J. Chem. Phys.* **92**, 6333 (1990).
[8] R. B. Metz *et al.*, *J. Chem. Phys.* **99**, 1744 (1993).
[9] M. J. Bronkowski *et al.*, *J. Chem. Phys.* **95**, 8647 (1991).
[10] S. S. Brown, H. L. Berghout, and F. F. Crim, *J. Chem. Phys.* **102**, 8440 (1995).
[11] N. Bloembergen and E. Yablonovitch, *Phys. Today* **31**, 23 (May 1978).
[12] F. F. Crim, *Science* **249**, 1387 (September 21, 1990).
[13] M. J. Coggiola *et al.*, *Phys. Rev. Lett.* **38**, 17 (1977).
[14] G. Porter and M. R. Topp, *Proc. Roy. Soc.* **A315**, 163 (1970).
[15] J, R, Novak and M. W. Windsor, *Proc. Roy. Soc.* **A308**, 94 (1968).
[16] R. L. Vander Wal and F. F. Crim, *J. Phys. Chem.* **93**, 5331 (1989).
[17] J. A. Paisner, *Appl. Phys. B* **46**, 253 (1988).
[18] I. L. Bass *et al.*, *Appl. Opt.* **31**, 6993 (1992).
[19] J. L. Lyman *et al.*, *Appl. Phys. Lett.* **27**, 87 (1975).
[20] S. R. Leone and C. B. Moore, *Phys. Rev. Lett.* **33**, 269 (1974).
[21] A. F. Bernhardt *et al.*, *Appl. Phys. Lett.* **25**, 617 (1974).
[22] F. Magnotta and I. P. Herman, *J. Chem. Phys.* **81**, 2363 (1984).
[23] K. Tamura, M. Oba, and T. Arisawa, *Appl. Opt.* **32**, 987 (1993).

Selected Additional References

D. L. Andrews, *Lasers in Chemistry*, Springer-Verlag, Berlin, 1990.
H. J. Arnikar, *Isotopes in the Atomic Age*, Wiley Eastern Limited, New Dehli, 1989.
M. A. El-Sayed, ed., *Laser Applications to Chemical Dynamics*, SPIE Proc., Vol. 742, SPIE, Bellingham, WA, 1987.
D. K. Evans, ed., *Laser Applications in Physical Chemistry*, Marcel Dekker, New York, 1989.
J. W. Hepburn, ed., *Laser Techniques for State-Selected and State-to-State Chemistry III*, SPIE Proc., Vol. 2548, SPIE, Bellingham, WA, 1995.
J. B. Jeffries, J. M. Ramsey, and R. P. Lucht, Laser Applications to Chemical Analysis, *Appl. Opt.* **34**, 3202 (1995), and following articles.
J. A. Paisner, ed., *Laser Isotope Separation*, SPIE Proc., Vol. 1859, SPIE, Bellingham, WA, 1993.
R. N. Zare, Laser Separation of Isotopes, *Sci. Am.*, p. 86 (February 1977).

Chapter 22 | Fiber Optics

Fiber optic cables are commonly used along with lasers to transmit laser beams to remote locations. They are perhaps best known because of their use in lightwave communications, but they have been employed for many other applications, such as beam delivery for material processing and as sensors.

Fiber optic elements are a type of optical waveguide. An optical waveguide, in its simplest form, is a small volume of transparent material with index of refraction higher than that of the surrounding material. Light inside the higher-index material tends to remain trapped there by the phenomenon of total internal reflection. If the material is in the form of a long (meters to kilometers), narrow (micrometers), flexible filament, it becomes an optical fiber element. Light inserted into one end of the fiber will be transmitted along the length of the fiber with low loss (at least in some cases) and emerge from the other end. The high-index core material has a cladding material covering it. The cladding has index of refraction lower than the core.

The highest-performance fibers utilize a glass core with a glass cladding. Fibers with a glass core and plastic cladding and fibers with plastic core and plastic cladding are also available. These options offer lowered cost with decreased performance.

Fiber optic devices have long been used to transmit light over short distances to inaccessible places. In the pre-laser era, they were utilized in many types of medical instrumentation designed for viewing internal body structures. The losses in the fibers were relatively high, so that their use was restricted to relatively short distances. In recent decades, the loses have been dramatically reduced, so that transmission over distances of many kilometers is feasible.

A. Structures

The structure of a glass fiber is shown schematically in Figure 22-1. The core, with index n_2, is surrounded by a cladding with index $n_1 < n_2$. In principle, the cladding

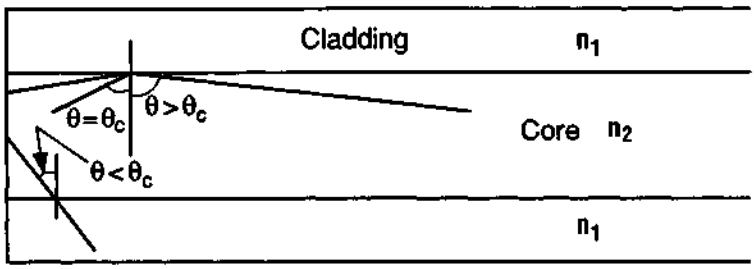


Figure 22-1 Fiber optic waveguide. The fiber, of circular cross section, has a core with index of refraction n_2 and a cladding with lower index n_1. Rays traveling at an angle θ (measured from the normal to the core–cladding interface) greater than the critical angle θ_c [θ_c = arc sin(n_1/n_2)] are confined by total internal reflection and propagate along the length of the fiber. Rays traveling at an angle θ less than θ_c may escape.

could be omitted, and one would have a waveguide due to the index difference between the glass and the surrounding air ($n = 1$). In practice, small imperfections on the outer surface of the glass would lead to high light losses, and so the cladding is always used in real devices. The thin glass fibers are flexible and can be formed in very long lengths by processes like drawing.

The loss of a fiber is commonly expressed in decibels per kilometer (dB/km). The use of this unit implies transmission over large distances. Before 1970, the best optical fibers had losses in excess of 1000 dB/km. Applications involving transmission over distances around 1 m, such as medical instrumentation, were feasible. But applications requiring fiber optic transmission of light over kilometer distances, like optical telecommunication, were not possible. Developments in glass technology in the 1970s steadily reduced the loss at wavelengths accessible to semiconductor diode lasers. By the mid 1970s, glass fibers with loss less than 1 dB/km were fabricated. Currently, fibers with loss of a few tenths of a dB/km at wavelengths of interest are available.

Simultaneously, fiber drawing technology was developed that allowed the production of continuous fibers with lengths of many kilometers. These advances enabled development of fiber optic telecommunication.

There exists a large body of theoretical work on the propagation of electromagnetic waves in waveguides. We shall not discuss the mathematics in detail here. We simply note that waveguiding can be understood in terms of total internal reflection. When light propagates across an interface between two media with different indices of refraction, n_1 and n_2, the angles θ_1 and θ_2 between the directions of propagation and the normal to the interface are related by Snell's law (which is discussed in elementary optics texts):

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \tag{22.1}$$

If we consider propagation from medium 2 into medium 1, with $n_1 < n_2$, we see that the maximum possible value of sin θ_2 is

$$\sin \theta_2 = \frac{n_1}{n_2} \tag{22.2}$$

Thus, light rays traveling in medium 2 at an angle θ_2 such that

$$\theta_2 > \text{arc sin} \frac{n_1}{n_2} \tag{22.3}$$

cannot escape from the medium. Such rays are totally internally reflected inside medium 2 and do not emerge. This is the basis of optical waveguiding. The minimum angle at which total internal reflection occurs is called the critical angle; its value is denoted by

$$\theta_c = \text{arc sin} \frac{n_1}{n_2} \tag{22.4}$$

The situation is sketched in Figure 22-1. Rays traveling at angles greater than θ_c are confined to the core of the fiber.

Optical fibers are characterized as single-mode fibers or multimode fibers. A mode is an allowed solution to Maxwell's equations for the particular fiber geometry. It represents in simple terms a path that a light ray may take as it reflects off the core–cladding interface during its propagation down the length of the fiber. Fibers with narrow cores may support only a single mode, whereas larger-diameter cores may support hundreds or thousands of modes. The condition for a fiber of diameter d to support only a single mode at wavelength λ is

$$\frac{\pi d(n_2^2 - n_1^2)}{\lambda} < 2.405 \tag{22.5}$$

For many typical cases, this occurs when the diameter of the core is less than a few micrometers. The number of modes present increases as the square of the diameter of the core.

Several different structures for fibers are illustrated in Figure 22-2. Multimode fibers offer ease of coupling light into the fiber because of the relatively large core. Multimode fibers have relatively large dispersion because of the variety of optical paths that a light ray may take as it travels through the fiber. Thus, a short pulse of light injected into one end of the fiber will undergo a temporal spread as it propagates. This limits the bandwidth of systems based on multimode fibers. Single-mode fibers have small dispersion because light can propagate only at small angles to the axis. Thus, there is less spread in the possible path lengths down the fiber and relatively little distortion of short pulses. Such fibers can be used in systems with large bandwidth. They require greater precision in fabrication, and it is more difficult to couple light into them efficiently.

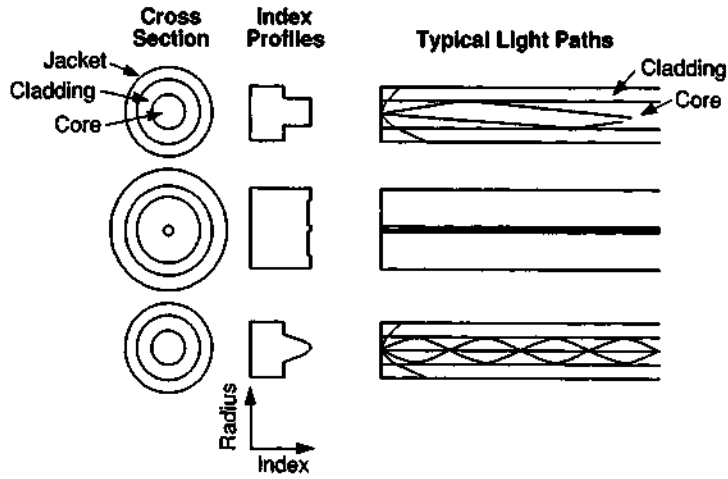


Figure 22-2 Types of optical fibers, showing cross sections, profiles of index of refraction, and paths of typical light rays. Top: Multimode fiber. Middle: Single-mode fiber. Bottom: Graded index fiber.

Fibers with graded index of refraction offer a compromise. A parabolic radial variation of the index of refraction acts like a continuous focusing element. Rays traveling at an angle to the axis are bent back toward the axis by the gradient in refractive index. A ray traveling at an angle to the axis travels farther than one along the axis, but it spends more time in regions with low index of refraction. Thus, it has higher average velocity than the axial ray. The choice of the parabolic profile tends to keep the different rays in phase and thus keeps dispersion small. The fiber is a multimode fiber, but short pulses of light are lengthened much less than in multimode fibers with step profiles of index. Because of the relatively large core, coupling of light into the fiber is relatively easy.

In some cases, it is desirable for the state of polarization of light launched into the fiber to be preserved. Single-mode fibers that preserve stable linear polarization of light are available. Single-mode fibers having very small cores with an elliptical cross section will preserve the polarization of light launched with its polarization vector parallel to the major axis of the ellipse. But it is difficult to couple light into these fibers because of the small core dimension. Polarization-preserving fibers with larger cores have been developed by fabricating the fibers with patterns of internal stress. The stress induces birefringence in the fiber. Light launched into the fiber with the polarization properly aligned will have its polarization direction preserved.

B. Losses

One of the most important characteristics of optical fibers is their attenuation. The attenuation for germania-doped silica core fibers as a function of photon energy is

shown in Figure 22-3. This curve is typical of the characteristics of many types of fibers. At relatively large photon energy (short wavelength), the loss is dominated by Rayleigh scattering, which increases as the inverse fourth power of the wavelength. In the lower-energy (longer-wavelength) region, the loss arises from the tail of the infrared absorption, dominated by multiphonon scattering. These two phenomena are represented by the dashed curves, which represent the minimum possible losses in these regions. The actual loss values of modern fibers are in fact fairly close to these limiting values. In the region near 1 eV photon energy (near infrared), the loss is at a minimum. At wavelengths near 1550 nm, the attenuation reaches its lowest value of a few tenths of a dB/km. The local peaks near 1 μm are caused by impurity absorption. The reduction in fiber loss over the last two and a half decades has largely been a matter of reducing impurity levels in the fibers.

In Chapter 24 we will show an expanded view of the loss as a function of wavelength in the region near 1 μm. This is the region important for lightwave communications.

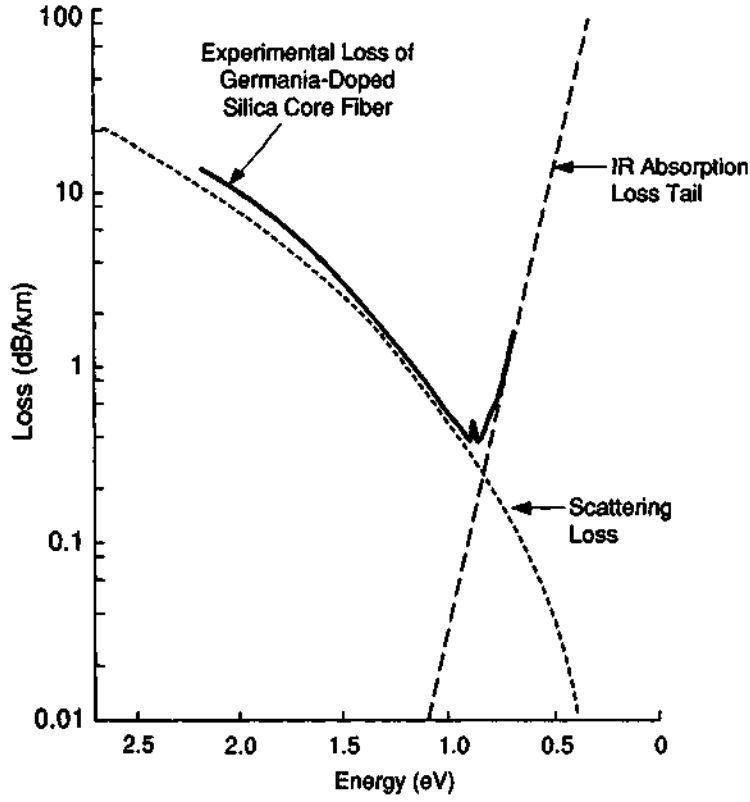


Figure 22-3 Loss of a germania-doped silica fiber as a function of photon energy.

Historically, as the purity of fibers was increased and their losses reduced, there was a local minimum in the attenuation near 850 nm, in a region accessible to gallium arsenide and aluminum gallium arsenide lasers. The first fiber optic telecommunication systems were developed using these laser sources and operating near this wavelength.

As the purity increased further, another local minimum developed near 1300 nm. This wavelength region was accessible to indium gallium arsenide phosphide laser sources, and telecommunication systems using these lasers at this wavelength were developed. The nomenclature employed was "short wavelength" for the 850 nm systems and "long wavelength" for the 1300 nm systems. This nomenclature was used during the period when the fibers were still developing. It may be found in older descriptions of optical communication links.

As the purity was increased still further, the lowest-loss region moved to 1550 nm, as Figure 22-3 indicates. This region, also accessible to indium gallium arsenide phosphide lasers, is now most often used for modern fiber optic lightwave communication systems. Nomenclature used when this region was being developed was "first window" for 850 nm, "second window" for 1300 nm, and "third window" for 1550 nm.

These advances, which required many years and intensive development efforts, were driven by the needs of lightwave communications. They led to the availability of low-loss optical fiber suitable for long distance communications and have led to the replacement of copper wire with glass fiber for many communications applications. These applications will be discussed further in Chapter 24.

The losses described here have been those associated with absorption by impurities, which have now been reduced to a very low level, and to fundamental physical phenomena, like multiphonon absorption. There are several causes for additional loss if the fiber is not fabricated well. Figure 22-4 shows, in an exaggerated manner, additional possible sources of loss if the fiber is less than perfect. Scattering from local imperfections in the core and loss due to microbending of the surface between core and cladding can increase the attenuation to higher levels than the absorption due to impurities.

So far, we have discussed losses in a straight fiber. Bending of the fiber may also introduce loss. There is a minimum bend radius, below which there is excessive loss because of radiation of light out of the fiber. A dielectric waveguide will radiate in regions where there is bending of the guide. As an example, for single-mode fiber formed with 1 percent difference of index of refraction, the radiation loss is negligible for bend radii greater than 1 mm. For a bend radius of 0.5 mm, the radiation loss becomes intolerable [1]. Bending loss can be a problem in cases where fibers are free to flex.

Light is coupled into the fiber through the end of the fiber. The light source, usually a laser or light emitting diode, is positioned adjacent to the end of the fiber, with the emitting area lined up with the core of the fiber. In practice, it is difficult to couple all the available light into the fiber. The diameter of the core may be only a few micrometers in diameter, so that alignment of the source and the core is critical. As the diameter of the core increases, this problem is reduced.

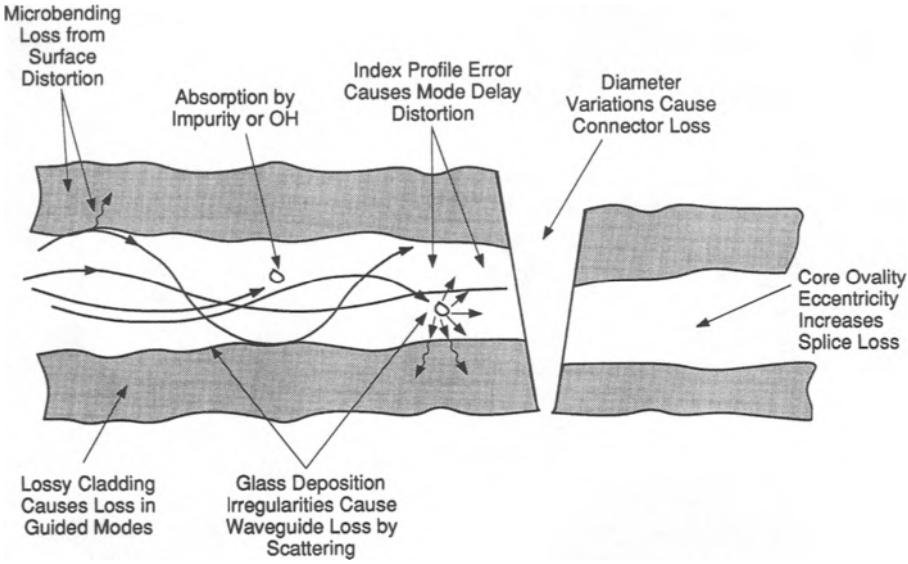


Figure 22-4 Structural and compositional defects leading to loss in optical fibers. (From *Amphenol Fiber Optic Designer's Handbook*, Amphenol Corp., Lisle, IL, 1987.)

There are several causes for loss of light from the source that does not get coupled into the fiber:

Reflection loss There is some loss because of reflection of light from the end of the fiber. This may be reduced to a low level if the end of the fiber is anti-reflection coated.

Unintercepted light loss This loss occurs when some of the light from the source does not fall on the fiber core. It arises if the diameter of the light source is larger than the diameter of the core, or if the source is not lined up properly with the fiber axis.

Nonconfinement loss Even if the light enters the fiber core, it will not be confined if it is traveling at a large angle to the fiber axis. There is a maximum acceptance angle within which incident light rays must strike the fiber so as to be trapped within the core. Losses occur if the beam divergence of the source is large, or if the axis of the emission profile does not line up with the fiber axis.

The maximum acceptance angle, α_m, defined from the normal to the end of the fiber, is given by

$$\sin \alpha_m = (n_2^2 - n_1^2)^{1/2} = \text{NA} \tag{22.6}$$

The parameter NA will be defined presently.

Figure 22-5 illustrates the situation. If the incident light falls within a cone of half-angle α_m, it will strike the walls of the fiber at an angle greater than the critical angle and be confined within the fiber.

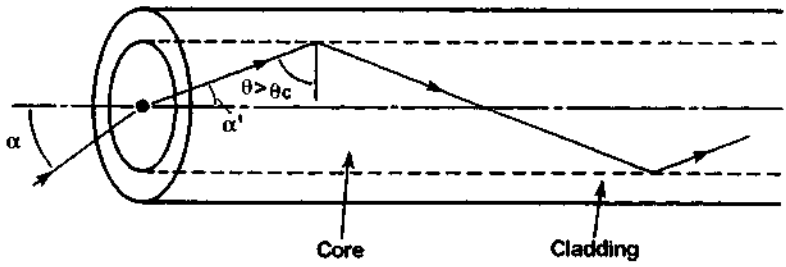


Figure 22-5 Acceptance angle for a fiber. If the angle α is less than the maximum acceptance angle α_m defined in the text, the incident light will be confined within the fiber. The angle θ is greater than the critical angle for total internal reflection.

Equation (22.6) also defines a quantity called the numerical aperture (NA), which is a measure of the light gathering power of the fiber. The numerical aperture of the fiber, as a measure of how well the fiber accepts light, is an important specification for optical fibers. The numerical aperture of single-mode fibers is typically around 0.11; that of multimode fibers is often in the range 0.2–0.3.

C. Manufacture of Optical Fibers

Optical fibers are prepared by pulling from the heated tips of preforms in the shape of solid rods, typically several millimeters in diameter and several centimeters long. The preforms may be prepared by a number of techniques, but probably the most common is chemical vapor deposition. In one process, the preform is fabricated from a very pure silica tube. Gaseous reactants, such as O_2, $SiCl_4$, and PCl_3, flow through the heated tube and deposit many thin layers of material on the inside of the tube. Then, the tube is heated to 2000°C and shrunk to form the solid preform.

In the preparation of the preform, the radial variation of material composition is controlled in a defined fashion. The radial composition of the preform is preserved in the fiber, so as to define the desired variations in index of refraction across the diameter of the fiber. The impurities are controlled to a very low level.

The preform is positioned vertically in a drawing tower. The tip is heated in a furnace, and the fiber is pulled from the softened tip. Polymeric coatings are applied to the fiber immediately after drawing, to protect it. The coated fiber is wound on a spool. A preform may be drawn into a roll of fiber many kilometers long.

The drawn fiber is then packaged to form a cable. The packaging process adds outer protective structures to the fiber. In many cases, a number of fibers may be packaged together in a multifiber cable. Many cables contain the following components:

Coated optical fiber or fibers

Buffer layer This is a plastic coating or plastic tube surrounding the fiber. The purpose of the buffer is to isolate the fiber from mechanical forces.

Strength members These materials, which include Kevlar, steel, and fiberglass, add mechanical strength to the cable.

Jacket The outer jacket, often black polyurethane, provides protection from the environment. The material of the jacket will vary, depending on the specific environment in which the fiber will be used.

Thus, the final optical cable contains several concentric layers of jacketing and protective materials around the glass fiber.

D. Connectors, Splicing, and Couplers

In many applications, optical fibers are used in conjunction with connectors to build up a light transmission system. The function of a connector is to position two segments of optical fiber end to end so that light may pass from one fiber into the next. The use of connectors allows one to disconnect and reconfigure the fibers. Because the micrometer-sized cores of the two fibers must be abutted accurately, a fiber optic connector must provide precision alignment.

The use of a connector involves some loss of optical power, the so-called insertion loss. The insertion loss of a connector should be as low as possible. Even very small misalignment will cause significant loss of transmitted light. Misalignments include

Lateral misalignment The centers of the cores are displaced.

End separation There is a gap between the ends of the fibers.

Angular misalignment There is a nonzero angle between the axes of the fibers.

The losses arising from these three causes is illustrated for multimode fiber in Figure 22-6. Even slight amounts of misalignment can lead to severe degradation of the transmitted optical power between the two fibers.

A variety of types of connectors have been devised to join fibers. They are used to connect fibers to transmitters, amplifiers, or receivers. Different models use ceramic ferrules, V-grooves, or aligned plugs and alignment sleeves. As one example, Figure 22-7 shows a diagram of a so-called biconic connector, popular for telephone applications. This connector uses a biconical alignment sleeve that aligns the axes of the two fibers. This type of connector is used for both multimode and single-mode fibers.

Commercial connectors are available that provide connection loss of a few tenths of a dB, and for which the loss does not increase after hundreds of connections and disconnections. The designer of a fiber optic system must choose a connector from among the many types available in an exercise involving trade-offs between the connector insertion loss, its cost, the ease of installation, and the repeatability of the loss as the connector is connected and disconnected many times.

In addition to connectors, splicing technology is required for field service and repair applications. Splicing of fibers may be accomplished in several ways. One

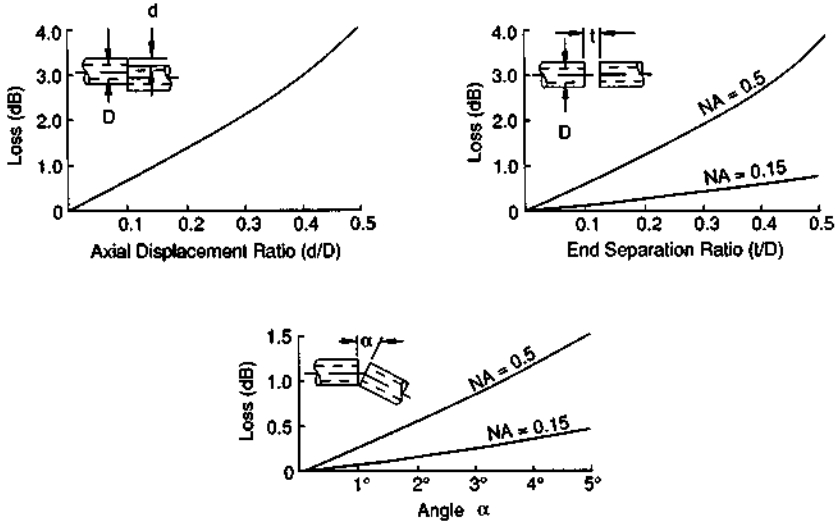


Figure 22-6 Losses arising from mechanical misalignment of two optical fibers for lateral misalignment (top left), end separation (top right), and angular misalignment (bottom). For the last two cases, results are given for two different values of the numerical aperture (NA). The nature of the misalignments and the notation are given in the insets.

method uses mechanical V-groove splice connectors to align the fiber ends. The ends are then bonded with index-matching epoxy adhesive. This method is relatively easy and inexpensive. It produces splices with losses around 0.2 dB. A second method uses an electric arc to weld the fiber ends together. The fiber ends are computer aligned by the splicing equipment, which is sophisticated and expensive. This method produces splices with loss as low as 0.05 dB.

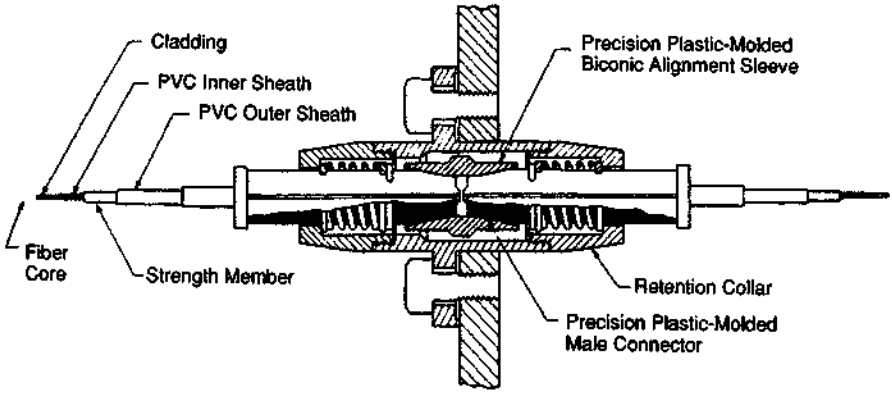


Figure 22-7 Biconic connector.

Another component commonly employed in fiber applications is a coupler, which is a multiport device used to distribute optical power among fibers. Light traveling in an optical fiber is not completely confined to the core; it extends somewhat into the surrounding cladding. Thus, if the cores of two fibers are brought close together, some light energy will transfer from one fiber to the other.

Couplers can take many forms. In one common adaptation, a fiber is glued into a groove in a quartz block. The fiber and block are polished to within a few micrometers of the core. Two such polished fibers are placed with their polished edges in contact. Then light propagating in one fiber will couple partially into the other.

Two common manifestations of couplers are tee couplers and star couplers. Tee couplers are three-port devices, which allow splitting off of a portion of the light in one fiber and transferring it to a second fiber. Star couplers are multiport devices that allow the distribution of light from any one port equally to all other ports.

Couplers are employed in many fiber optic applications. In particular, they are used in fiber optic sensors and in short-haul communication systems, such as local area networks.

E. Fiber Amplifiers

A relatively recent advance in fiber optics is the development of the erbium-doped fiber amplifier (EDFA). A length of fiber with the element erbium added can act as an amplifier for light in the 1550 nm wavelength region. The EDFA is pumped with light at a wavelength of 980 nm. As the 1550 nm light passes through the EDFA, its power is increased by transfer of energy from the pumping light.

In long distance lightwave communication links, the signal must periodically be detected, amplified electronically, and retransmitted in order to keep the signal at an acceptably high level. The significance of the EDFA is that it offers an all-optical means for amplifying the optical signal. There is considerable interest in using erbium-doped fibers to amplify weak signals in lightwave communication systems.

So far, we have emphasized the capabilities of fibers operating in the near infrared portion of the spectrum, suitable for communication applications. We turn now to two other topics: fibers for use in the longer wavelength infrared, and the potential use of optical fibers as the basis for sensor instrumentation.

F. Infrared-Transmitting Fibers

Silica-based fibers have high attenuation at wavelengths longer than about 2 μm. It would be desirable to have optical fibers for applications at longer infrared wavelengths. Such fibers could be used in medical applications and in material processing to deliver CO_2 laser radiation to workpieces. These fibers have higher attenuation than the well-developed silica-based fibers; attenuation coefficients are measured in dB/m rather than dB/km. Three main types of infrared-transmitting fibers have been developed, plus hollow waveguide structures.

Fluoride glasses These materials are based on fluorides of heavy metal elements, such as zirconium and barium. One example is the so-called ZBLAN fiber, which incorporates zirconium, barium, lanthanum, aluminum, and sodium. Such fibers are useful over the wavelength range 0.4–5 μm. This covers the wavelength of the Er:YAG laser, near 2.9 μm, which is being considered for medical applications. The minimum attenuation of these fibers, which are commercially available, is in the range 0.015–0.025 dB/m, in the wavelength region from 1.5–2.75 μm.

Chalcogenide glasses These glasses incorporate metals like arsenic, germanium, and antimony along with heavier elements from column VI of the periodic table (sulfur, selenium, and tellurium). One of the most widely used combinations employs germanium, arsenic, and selenium. They are useful over the wavelength range 2–11 μm, a range that covers the CO_2 laser emission. The chalcogenide glasses are one of the very few types of glassy materials that transmit in the long wavelength infrared. These fibers also are commercially available. Their minimum attenuation is around 0.7 dB/m in the 5 μm region. Near 10 μm, the attenuation is typically several dB/m; thus, although they can transmit CO_2 laser light, there is substantial loss if the path is long. Because of the poisonous nature of some of the constituents, these glasses are not good candidates for medical applications involving insertion into the human body. Chalcogenide fibers have demonstrated the ability to transmit substantial amounts of laser power. Arsenic sulfide and germanium arsenic sulfide fibers have transmitted multihundred watt power levels from a CO laser at a wavelength near 5 μm [2].

Crystalline fibers A number of crystals are transparent in the infrared and can be extruded as polycrystalline fibers. These materials include halides of thallium, silver, and alkali metals like potassium. Materials that have been used include TlBr, TlCl, AgCl, AgBr, KCl, and KBr. Usually alloys like $AgCl_xBr_{1-x}$ are used instead of the simple compounds. These materials have been extruded as fibers, which can cover the wavelength range from 2 to 20 μm. In theory, such materials could have very low attenuation, as low as 10^{-5} dB/m; in practice, the minimum attenuation coefficients that have been demonstrated are around 1 dB/m. Fibers of these materials are very soft and thus are hard to use in practical applications. They have demonstrated the ability to deliver substantial amounts of CO_2 laser power. One-meter-long thallium halide fibers can transmit nearly 100 W of continuous CO_2 laser radiation. Despite extensive research work, these fibers seem not to have reached widespread commercial availability.

Hollow metal waveguides Although these devices are not fibers, they are included here because they compete for the same applications as infrared-transmitting fibers. These waveguides are fabricated from materials like aluminum strips. Alternatively, they have been formed by vacuum deposition of metals like silver. Such devices can have high transmission (>90 percent for 1 m lengths) and higher power handling capabilities than fiber materials. Transmission of more than 500 W of continuous CO_2 laser power has been demonstrated. These devices have been the subject of substantial development and have been used in specialized applications, but they are not widely available commercially.

Despite long-standing research and development efforts on several material combinations, the losses of fibers for use in the longer wavelength infrared region remain higher than desired, and their applications have been slow to develop.

G. Fiber Optic Sensors

In addition to their use as transmission media, optical fibers have many potential applications as sensors. Sensor applications exploit the fact that the effect of an environmental parameter can change the propagation of light in a fiber. Optical fiber sensors can detect many physical and chemical quantities, including temperature, pressure, vibration, flow, liquid level, proximity, acceleration, angular rotation, magnetic field, acoustic waves, and chemical concentrations. A basic fiber sensor consists of a light source (usually a laser or LED), a length of fiber, and an optical detector. The parameter being sensed acts on the fiber and changes the amount of light reaching the detector. A variety of possible sensing mechanisms are employed in fiber sensors. It is not possible to review all of them here, but we shall present a few representative examples.

Extrinsic sensors The sensing effect takes place outside the fiber. An example would be a chemical sensor that utilizes a chemically sensitive material on the tip of the fiber. The fiber tip is in contact with a chemical solution. Light is launched into the fiber and reflected back through the fiber. As the concentration of the chemical in the solution changes, the properties of the material on the tip change, and the amount of light reflected back changes. The change gives a measure of the chemical concentration.

Intrinsic sensors Intrinsic sensors utilize a change that takes place within the fiber itself. This is in contrast to extrinsic sensors, in which the change is outside the fiber and the fiber itself remains unchanged. An example is a pressure sensor based on microbending of the fiber. A fiber is placed between two plates with rough surfaces. As pressure increases, the plates press on the fiber and induce many microbends in it. This increases the attenuation of the fiber. The power transmitted through the fiber decreases, and the magnitude of the decrease gives a measure of the pressure.

Interferometric sensors These sensors utilize two fibers. Light is transmitted through both fibers. One fiber (the reference arm) is isolated from the environment, and its properties remain constant. The other fiber (the measurement arm) is exposed to the parameter being sensed. The parameter changes the phase of the light being transmitted through it. For example, as a strain sensor, the length of the optical path in the fiber would change. The two beams are recombined after they emerge from the fibers to form an interference pattern. The accumulated path difference in the two arms can be derived from the interference pattern, and this can give a measure of the strain. Figure 22-8 shows a schematic diagram of the operation of an interferometric fiber sensor. Interferometric sensors offer the greatest sensitivity and potentially the highest performance capabilities for fiber sensors.

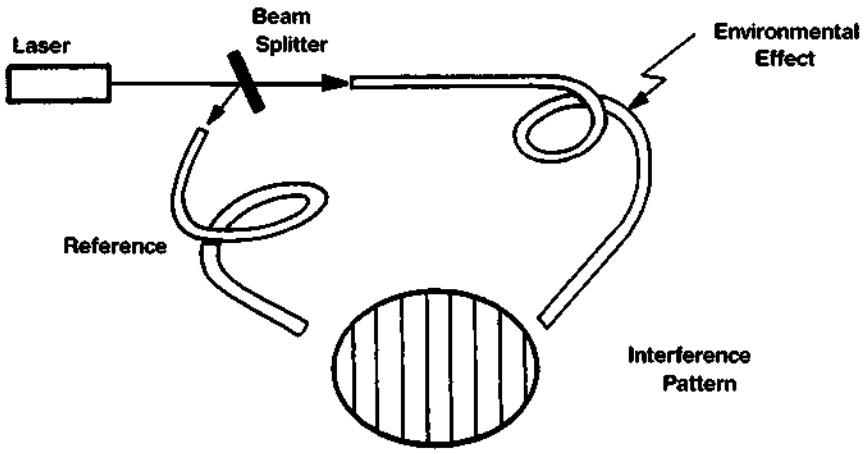


Figure 22-8 Interferometric fiber optic sensor.

Fiber optic sensors have been the subject of extensive research and development for well over a decade. The operation of fiber sensors to monitor a large number of different parameters has been demonstrated. Perhaps the most intensive development has been on fiber optic gyroscopes (FOGs), which were described briefly in Chapter 11. Fiber optic gyros are useful for applications like attitude and heading reference systems and should reach inertial navigation status in the near future. Another application that has undergone extensive development is acoustic hydrophone arrays.

Fiber optic sensors offer several advantages, including small size and light weight, immunity to electromagnetic interference, potentially high sensitivity, completely passive operation, and environmental ruggedness. A number of simple fiber optic sensors were developed early and were rapidly brought to commercial status. Some of the more complex systems, especially interferometric sensors, like the FOG, have required longer development, and their introduction to commercial markets has been slower. The availability of fiber optic sensor products and their market penetration continues to increase.

We present a single example of a fiber sensor developed for environmental monitoring applications [3]. The fiber is intended to detect leaks from fuel tanks that store petroleum products. Such leaks can contaminate soil and groundwater in the vicinity of the tanks. Conventionally, monitoring has been done by taking samples of soil from the area around the tanks and performing chemical analysis. The availability of an in situ real-time sensor offers many advantages.

The sensor uses an optical fiber from which the cladding has been removed and replaced with a coating that absorbs hydrocarbons. When molecules of hydrocarbons are absorbed, they change the index of refraction of the coating. This in turn changes the loss of the fiber. Thus, the presence of hydrocarbons causes the light intensity transmitted through the fiber to change.

The sensor responds to materials like gasoline, diesel fuel, jet fuel, toluene, and benzene. These materials can be detected at levels below one part per million. The sensors may be installed in the area around and under fuel storage tanks, to provide fast and continuous monitoring of the level of hydrocarbons present and to serve as an indicator of leaks.

H. Summary

We conclude the chapter with a description of the advantages and applications of fiber optics.

As a carrier for lightwave communications, fiber optic cable offers many advantages:

- It has the capability for very large bandwidth and for very high data transmission rates.
- Compared with copper wire, it is substantially lighter and smaller in size. This makes it suitable for use in airborne applications. It may be used to replace copper in conduits to increase the data rate for existing data links.
- It is immune to electrical shorts, electromagnetic interference (EMI), electromagnetic pulses (EMP), and radio frequency interference (RFI).
- It is relatively secure. It is difficult to tap without detection.
- The signal is well confined within a fiber. This eliminates crosstalk.
- It has lower cost per installed voice circuit.

Because of these advantages, optical fibers have become widely used in many communications applications, and they have in fact revolutionized telecommunications. This will be discussed in more detail in Chapter 24.

In addition, optical fibers have become widely used for delivery of laser beams to inaccessible locations. Many laser-based material processing systems employ fiber optics for beam handling and for beam distribution to multiple processing locations.

Optical fiber multiplexing modules are commercially available. They are generally used with Nd:YAG lasers, because the laser wavelength is in a region of low fiber absorption. Commercial units can handle average laser power up to 400 W. Available devices allow splitting of a laser beam into several fibers for simultaneous delivery of energy from one laser to several different workstations. Alternatively, they may be used on a time sharing basis, with different fibers delivering the beam on demand to workstations performing different operations. The availability of such devices provides the user with great flexibility to use a laser conveniently for a variety of material processing tasks.

Finally, optical fibers provide the basis for an emerging class of sensor instrumentation.

References

[1] S. E. Miller, *IEEE J. Quantum Electron.* **QE-8**, 199 (1972).
[2] S. Sato, K. Igarashi, and M. Taniwaki, paper CWJ5 at the Conference on Lasers and Electro-Optics, Anaheim, CA, May 10–15, 1992.
[3] D. P. Saini and S. L. Coulter, *Photonics Spectra*, p. 91 (March 1996).

Selected Additional References

P. K. Cheo, *Fiber Optics: Devices and Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
S. L. W. Meardon, *The Elements of Fiber Optics*, Regents/Prentice Hall, Englewood Cliffs, NJ, 1993.
S. E. Miller and A. G. Chynoweth, *Optical Fiber Telecommunications*, Academic Press, San Diego, 1979.
S. E. Miller and I. P. Kaminow, *Optical Fiber Telecommunications II*, Academic Press, San Diego, 1988.
J. P. Powers, *An Introduction to Fiber Optic Systems*, Irwin, Homewood, IL, 1993.
E. Udd, ed., *Fiber Optic Sensors*, Wiley, New York, 1991.
C. Yeh, *Handbook of Fiber Optics: Theory and Applications*, Academic Press, San Diego, 1990.

Chapter 23 | Integrated Optics

The term "integrated optics" refers to a variety of techniques that have been employed to produce compact optical circuitry [1]. The techniques are analogous to those employed in integrated electronic circuitry. The components are very small and can be integrated on substrate materials in circuits that include sources, couplers, filters, detectors, and other active and passive components needed for optical signal transmission and processing. The structure can be packaged as a chip designed to perform a task. Transmission of light over a distance is carried out in an optical fiber mated to the chip. The techniques of fabrication are similar to those for integrated electronic circuitry, with the result that devices can be batch-produced economically.

Historically, optical components have been relatively large devices, with dimensions of the order of centimeters or greater. Typical circuitry might include such components as a laser, modulator, and a detector to form, for example, a communications system. The light beam would propagate freely through the atmosphere and could be recollimated periodically by means of lenses.

Integrated optics employs waveguide transmission of light. As the last chapter described, light can propagate in a medium that has higher index of refraction than the surrounding material. The propagation can have low loss. The ability of light to propagate as a trapped wave in a waveguide is one of the central foundations of integrated optics, just as it is for fiber optics. Other circuit devices, such as lasers, modulators, and detectors, can be fabricated in the same material as the waveguide and can couple directly into the waveguide. The dimensions of the waveguide and the associated devices can be very small, so that many devices can be packed closely. Thus, one is led to the idea of an electrooptic integrated circuit (EOIC), similar in concept to an electronic integrated circuit.

The approach of integrated optics leads to compact, rugged devices that may be easily and economically fabricated. Compared with using propagation of light through the atmosphere, integrated optics offers advantages of freedom from variations and losses in the atmosphere. It also offers savings in size and cost.

Much of the development of integrated optical devices has involved use of lithium niobate as the basic material. Lithium niobate offers excellent electrooptic and waveguiding properties. There have been many demonstrations of integrated optic structures based on lithium niobate.

But to integrate a source and detector as part of an integrated circuit, one needs semiconductor materials, like GaAs or GaAs/AlGaAs. With semiconductor materials, one can fabricate the entire circuit, including emitters, detectors, and controlling and guiding optics, plus electronic drive circuitry, on the same chip of material. Thus, this chapter will emphasize use of devices based on semiconductor materials.

Integrated optical circuitry is a rapidly expanding field, but it is one still largely in development. Integrated optics has not yet replaced all the larger individual lasers and circuit elements that have traditionally been used, although applications and commercial availability are growing.

In this chapter, we shall briefly survey some of the materials and techniques that are involved in integrated optics. This will not be an exhaustive discussion, even of important developments; rather, we will present a few examples of approaches that are being developed.

A. Optical Waveguides

As Chapter 22 described, an optical waveguide is a small volume of transparent dielectric material with index of refraction higher than that of the surrounding material. Figure 23-1 shows an elementary waveguide, with the index n_2 of the surface film greater than the index n_1 of the substrate. An optical wave, once trapped inside the medium with higher index, will propagate along the waveguide and be confined by total internal reflection. In the structure shown in the figure, light launched into the waveguide at the input position will be transmitted through the structure and emerge at the output position.

One can distinguish between two types of optical waveguide structures:

1. The medium in which the light propagates is different from the surrounding medium. In this case, the waveguide should be thin, preferably thinner than one wavelength of light. This is the case illustrated in Figure 23-1, in which the waveguide is a thin surface layer, with one side in contact with air. Such waveguides may be formed by sputtering thin films of materials like zinc oxide or tantalum oxide onto glass substrates.

   The structure in Figure 23-1 will act as a transmission line. The material in the guiding region may have an index of refraction 1 percent (or less) higher than that of the bulk material. Typical dimensions are of the order of a few micrometers. This leads to high values of irradiance in the waveguide, even with relatively low total transmitted power.
2. The second general class of optical waveguides involves an embedded guide with a refractive index that is only slightly higher than that of the surrounding

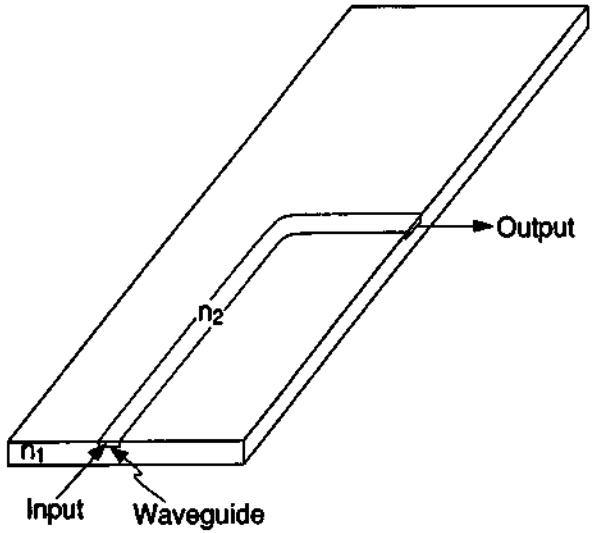


Figure 23-1 An optical waveguide in the form of a thin surface film. The index of refraction n_2 of the film is greater than the index of refraction n_1 of the substrate.

medium. Because the difference in index of refraction between the waveguide and the surrounding material is very small, the waveguide can have dimensions larger than the wavelength. For such optical waveguides, the index of refraction can be only slightly higher (~0.1 percent) than that of the surrounding medium. Such a structure is illustrated in Figure 23-2. This type of waveguide can be formed by techniques such as diffusion or ion implantation. Optical fibers can be considered as a type of embedded guide, with the core embedded in the cladding.

The structures shown in Figures 23-1 and 23-2 have a sharp discontinuity in index of refraction at the edge of the confining region. This is not necessary; the gradient of the index of refraction can be gradual and optical waveguide operation will still occur.

Surface waveguides may be fabricated in a number of ways:

Ion bombardment Ion bombardment can produce regions of high index of refraction at the surface of materials such as fused silica.

Ion exchange Diffusion methods have produced films with graded index of refraction on the surface of glass. For example, floating molten glass on the surface of molten tin yields so-called Pilkington float glass, which has a surface layer with higher index than the bulk material.

Sputtering Radio frequency sputtering of glass and of semiconductors will produce optical waveguide films.

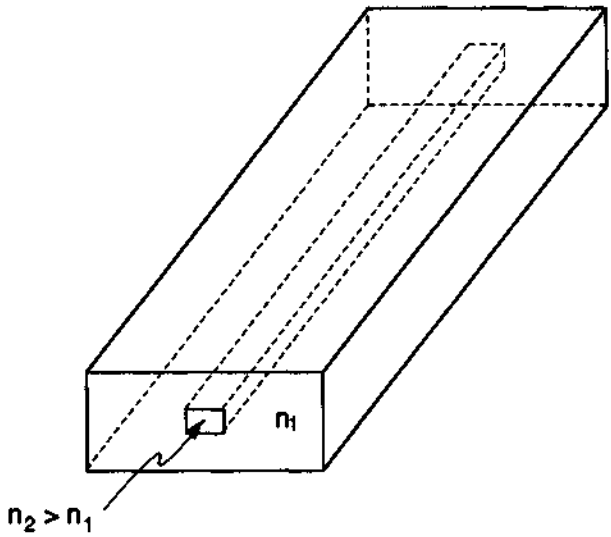


Figure 23-2 Optical waveguide (index of refraction n_2) embedded in a medium with index of refraction n_1.

Epitaxial growth techniques Epitaxial growth techniques offer a method of fabricating thin layers of material of variable composition and variable index of refraction on crystalline substrates. They are especially applicable to semiconductor materials, such as aluminum gallium arsenide. They are often used to fabricate multiple layers with different compositions. The methods include liquid phase epitaxy (LPE), in which a sliding assembly transports the substrate over a number of liquid melts containing material of the desired compositions; metal-organic chemical vapor deposition (MOCVD), in which gaseous materials, like $Ga(CH_3)_3$, containing the desired atomic species are brought to a heated substrate and react chemically to produce a layer of the desired composition; and molecular beam epitaxy (MBE), in which the constituent atoms are delivered to the substrate as molecular beams, derived from heated sources that may be shuttered to control the composition of the material. These methods all allow deposition of very thin layers of high-quality crystalline material of controlled composition. MBE especially yields very high purity films.

Once the surface films have been formed, it is necessary to use masking and etching techniques to form optical circuits. Conventional photolithographic techniques employ optically exposed photoresist and have been used to produce optical circuitry in surface films. Photolithography suffers because it produces an edge that is rougher than desired. Edge roughness leads to losses from the waveguide by radiation. Use of electron beams to expose photoresist has proved satisfactorily. This approach leads to high resolution and good edge definition.

Now we consider techniques for coupling of light into optical waveguides. Figure 23-3 shows three methods for coupling light from an external source into a thin film surface waveguide. These methods are

- Prism coupling
- Grating coupling
- Direct coupling

The same methods may be used to transform light propagating in an optical waveguide into freely propagating light outside the waveguide.

In the prism coupling method (Figure 23-3, upper left), frustrated total internal reflection at the base of the prism leads to excitation of a wave in the waveguide by a tunneling process. The gap between the prism and the film must be very small, less than one optical wavelength. The prism coupler is simple and versatile, and it can transfer more than 70 percent of the incident light into the film.

The grating coupler (Figure 23-3, upper right) employs frustrated total internal reflection of the diffracted first-order wave to excite the waveguide. The grating coupler is very adaptable in integrated optical circuitry and can be fabricated by techniques that are compatible with the rest of the device. The grating is rugged and is also capable of coupling more that 70 percent of a light beam into the film. Suitable gratings may be formed by photolithographic techniques, with exposure of photoresist to an interferometric fringe pattern.

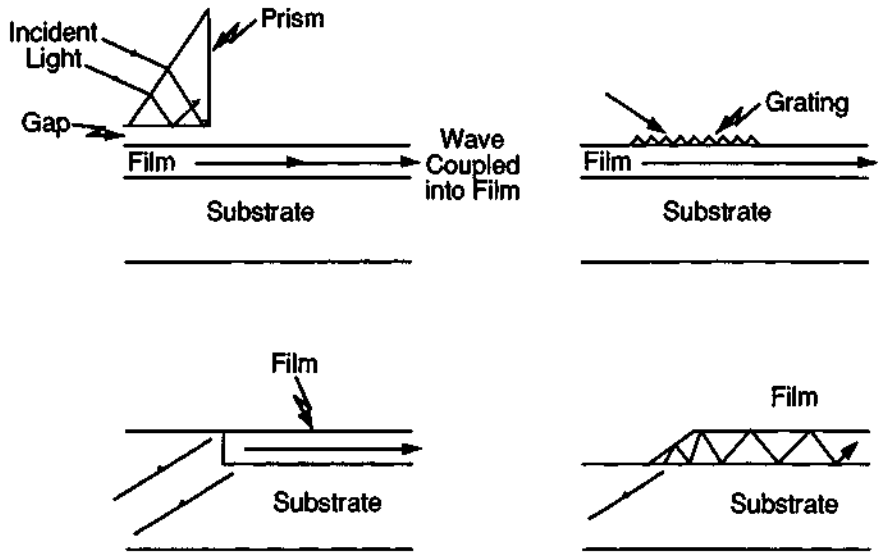


Figure 23-3 Methods for coupling light into a thin film surface waveguide. Upper left: Prism coupler. Upper right: Grating coupler. Lower left: Direct coupling, abruptly terminated waveguide. Lower right: Direct coupling, tapered waveguide.

The grating coupler and the prism coupler are most applicable when the waveguides are relatively wide. For narrow waveguides, these methods are less efficient, because the laser light has to be focused. In such cases, direct injection (Figure 23-3, lower left and lower right) is preferable [2]. The beam is inserted into the waveguide through the substrate by directing the beam at a small angle to the interface between the waveguide and air. The edge of the waveguide can terminate inside the substrate (Figure 23-3, lower left), or there may be a tapered edge (Figure 23-3, lower right). The direct injection method can couple more than 90 percent of the light into the guide.

B. Components for Integrated Optics

The point of integrated optics is to be able to form all the circuit components, for example, laser sources, couplers, modulators, and detectors, in thin film form on the surface of a substrate. The desired result is a complete packaged system comprising a number of components integrated on a chip. The fabrication processes must be compatible with large-scale economic production and with integration of the various components. There have been many demonstrations of individual components, but fewer examples of complete operating integrated optic circuits that have progressed to practical application.

We shall describe several examples of fabrication of circuit components, specifically modulators, thin film lasers, and deflectors. These examples are chosen to illustrate possible fabrication technology; they are far from an exhaustive list of possible devices or possible fabrication methods.

1. MODULATORS

As an example of how optical circuit elements can be fabricated in integrated optics, we discuss experiments regarding optical modulators using p–n junctions in semiconductors. This will illustrate some of the materials that can be used.

A schematic diagram of a semiconductor junction device developed as an optical modulator is shown in Figure 23-4. The p–n junction in the semiconductor acts as an optical waveguide. The waveguiding arises from the higher index of refraction in the plane of the junction. Confinement of light in a p–n junction has been known for many years. The operation of such waveguides is very efficient. Waveguides consist of thin epitaxial layers (approximately 10 μm thick) of semiconductors like gallium arsenide or gallium phosphide. These are deposited on substrates of the same semiconductor material with different resistivity. An alternative approach could use a junction in a film of GaAs (about 0.5 μm thick) with epitaxial layers of AlGaAs on both sides of the junction. The AlGaAs has lower index of refraction than the GaAs and provides good confinement of the light near the GaAs junction.

The depletion layer in such a junction is birefringent. The birefringence apparently arises because of electrooptic effects within the junction. In the figure, the applied voltage changes the birefringence, so that the structure acts as an optical

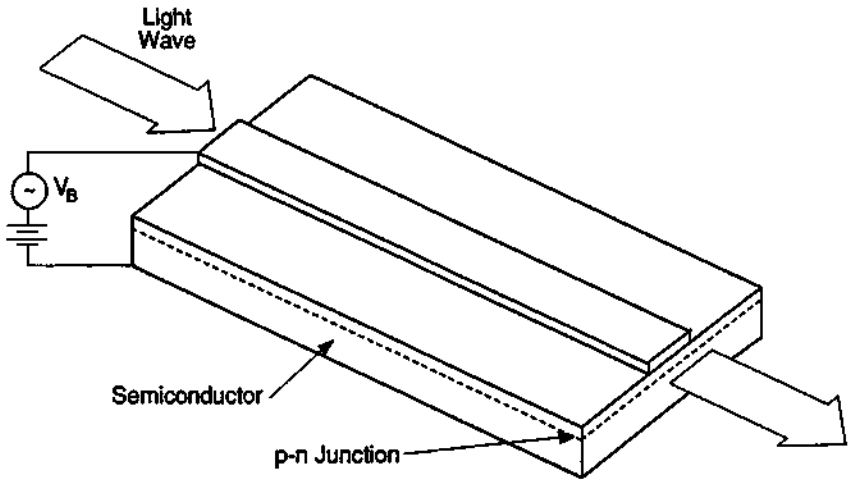


Figure 23-4 Optical modulator based on semiconductor junction. (From S. E. Miller, *IEEE J. Quantum Electron.* **QE-8**, 199 (1972).)

modulator. Because the light is well confined within the birefringent material, the operation of the modulator is efficient. For example, modulation of helium–neon laser light in semiconductor waveguide structures has been demonstrated with modest input power to the modulator and with wide bandwidth [3,4].

2. *DISTRIBUTED BRAGG REFLECTOR LASERS*

The techniques of integrated optics have been employed for producing lasers that consist of thin films on the surface of a substrate, and in which the feedback is provided by a grating rather than by a conventional mirror structure. The device is fabricated as an integrated structure including the laser and the grating on the same substrate.

One common form for such a device uses corrugated waveguide gratings. The grating structures are at the ends of the active laser medium. They act as mirrors by scattering light back into the active medium. Each bump in the corrugation reflects a small fraction of the light. At a wavelength λ satisfying the Bragg condition

$$\lambda = 2n\Lambda \tag{23.1}$$

where n is the index of refraction of the waveguide and Λ is the period of the grating, all the reflections are in phase and they interfere constructively. Such lasers are termed distributed Bragg reflection (DBR) lasers, DBR lasers were described briefly in Chapter 3. An example of a structure of a DBR laser is shown in Figure 23-5. In the device shown, the laser is a multiple quantum well. The three sources of current, which can be varied independently, are used to drive the laser, to change the length of the laser cavity, and to tune the wavelength of the Bragg reflector. The cavity

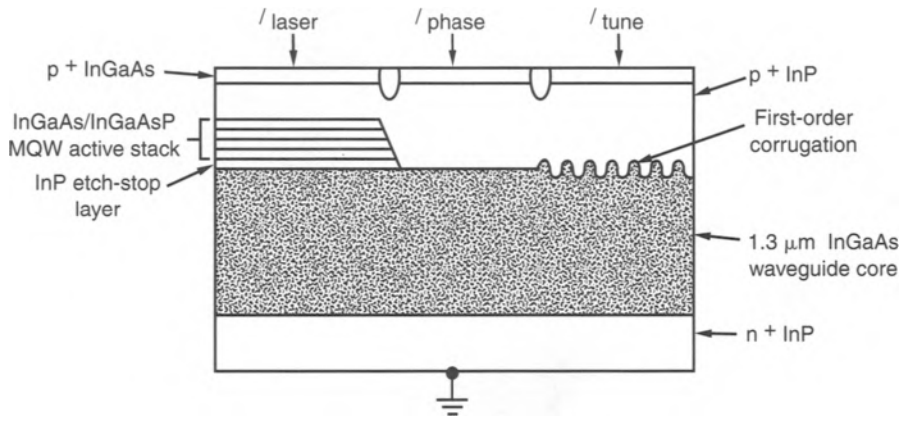


Figure 23-5 Example of an integrated distributed Bragg reflector laser. The three sources of current I indicated are for excitation of the laser (I_{laser}), changing the length of the laser cavity (I_{phase}), and tuning the Bragg reflector (I_{tune}). (From T. L. Koch and U. Koren, *AT&T Tech. J.* **71**, 63 (1992). Copyright © 1992 AT&T. All rights reserved. Reprinted with permission.)

length shift and the tuning occur because the index of refraction of the material varies as a function of current. These functions are used to tune the wavelength of the laser output.

The corrugations may be formed by allowing the beams from two ultraviolet lasers to produce an interference pattern in a layer of photoresist coated on the surface of a wafer. The photoresist is developed, and it forms a mask for etching the pattern of corrugations.

DBR lasers provide small, rugged structures that are well suited for integrated optics applications. Such thin film lasers are easily compatible with other integrated optic structures.

DBR devices can operate in a single longitudinal mode with narrow linewidth. Because of their narrow spectral width, they are commonly used as sources for lightwave communications.

3. ACOUSTOOPTIC DEFLECTORS

As a further example of how integrated structures can be formed to perform functions in optical circuitry, we consider an acoustooptic light beam deflector. The structure is shown in Figure 23-6. The waveguide had a thin glass film sputtered on the surface of an α-quartz crystal. A helium–neon laser beam was coupled into the waveguide by a grating coupler.

A surface acoustic wave was produced in the quartz crystal by an interdigital transducer. The acoustic wave set up a grating in the glass film. The optical wave was diffracted by the grating. The compressions and rarefactions associated with the acoustic wave acted as a diffraction grating for the light. Light was extracted at

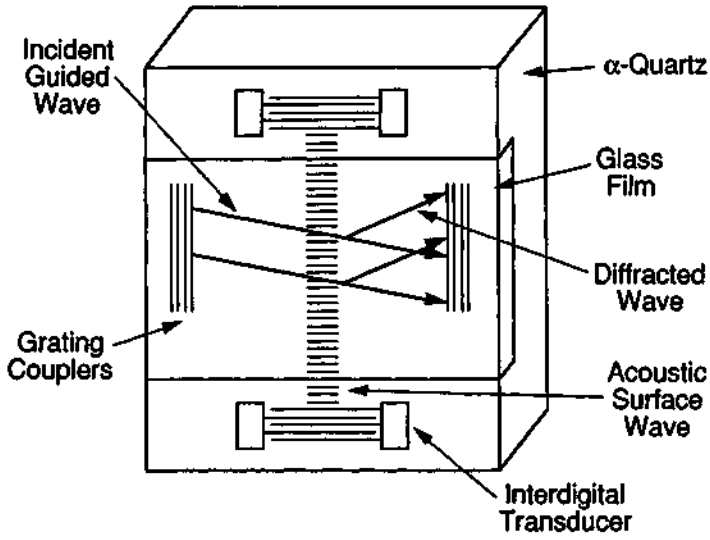


Figure 23-6 Thin film acoustooptic deflector. (From L. Kuhn *et al.*, *Appl. Phys. Lett.* **17**, 265 (1970).)

different angles, depending on whether the acoustic wave was present or not. The demonstration represented a merging of older technologies employed in acousto-optic deflection in larger structures with thin film optical waveguide techniques. This offers a good example of how different structures can be combined in an integrated optical circuit.

We have presented examples of how several components (laser, modulator, and deflector) can be formed in an integrated optic format. This by no means exhausts the list of components that have been demonstrated. We shall not attempt to describe all these different components nor the diverse methods of producing them. We shall simply allow these few components to serve as examples for the many that have been developed.

C. Integrated Optic Circuits

The use of integrated optical circuitry containing thin film waveguides with associated active and passive circuit elements is suitable for control of light beams. There are many advantages to this approach:

1. The circuitry is very small, compact, and lightweight; integration of a complete functionality on a tiny chip is possible.
2. Required circuit elements like sources, resonant cavities, directional couplers, filters, modulators, deflectors, and detectors can easily be fabricated as integral parts of a circuit.
3. Mature semiconductor integrated circuit processing and technology can be used to produce thin film optical waveguides and circuits. Well-developed

techniques like photolithography, sputtering, and ion implantation are used in the fabrication of devices. Complex circuit patterns may be formed using photolithographic techniques. The processing can thus be simple and inexpensive, and it can yield large quantities of devices.

4. Because the beam is confined to a small area, the irradiance is high. Nonlinear optical effects can easily be employed if nonlinear optical materials in thin film form are used.
5. If the film is an electrooptic material, a large electrooptic effect can be obtained with very low applied voltage. Beam modulation and deflection can be achieved with low power consumption.
6. The laser source and associated optical circuitry can be isolated from thermal fluctuations, mechanical vibrations, and airborne acoustic effects, which hinder the operation of separately mounted monolithic components.

Although integrated optics development is more than two decades old, it has not yet reached its full potential. Most practical applications of lasers still involve the use of relatively large, discrete components. But development of new materials, techniques, and components for integrated optics is proceeding rapidly. We will present a few examples, out of many possible, that illustrate complete integrated optic devices.

1. RADIO FREQUENCY SPECTRUM ANALYZERS

One of the first demonstrations of integrated optical circuitry involved radio frequency spectrum analyzers [5,6]. In one manifestation [5], a waveguide defined by titanium diffusion on a lithium niobate substrate had a diode laser beam butt coupled into it. A surface interdigital transducer array on the lithium niobate generated surface acoustic waves. The diode laser beam was collimated by one guided wave lens and focused onto a linear array of photodiode detectors by a second lens. The photodiode array, containing 140 elements, was butt coupled to the opposite side of the assembly from the diode laser.

The radio frequency signals were coupled to the transducer array. The laser beam was deflected by the surface acoustic waves generated by the transducers. The deflection of the beam was proportional to the frequency of the radio frequency signal. Thus, the frequency could be determined by the position of the laser beam at the photodiode array.

2. TRANSMITTERS

Applications in communications seem particularly favorable and provide the motivation for much of the interest in integrated optics. Integrated optic transmitters, repeaters, and receivers can offer wide-bandwidth systems that require smaller volume than conventional devices.

There have been a number of demonstrations of optical transmitters with the laser and the drive electronics fabricated on the same chip of semiconductor material. Figure 23-7 shows one example of a monolithic laser transmitter fabricated from

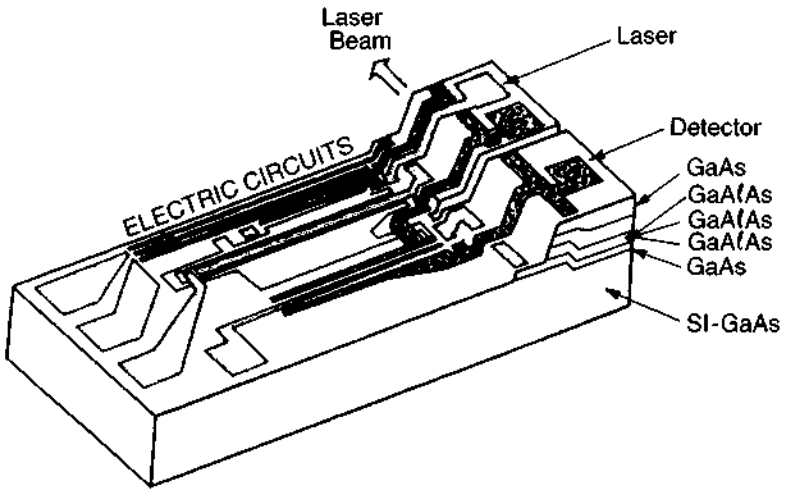


Figure 23-7 Monolithic integrated optic transmitter. (From H. Matsueda, S. I. Sasaki, and M. Nakamura, *J. Lightwave Tech.* **LT-1**, 261 (1983) © 1983 IEEE.)

GaAs/AlGaAs. The circuitry contains a laser with an on-chip mirror, the laser driver circuitry, and a photodiode with an amplifier, for monitoring purposes.

This example shows the integration of electronic and optical devices fabricated from the same materials. There are still advances required in the compatibility of the material for both types of functions.

3. SOURCES FOR WAVELENGTH DIVISION MULTIPLEXING

Figure 23-8 shows an integrated optic circuit designed as a source for a wavelength division multiplexed (WDM) system intended for lightwave communication applications. The source consists of four tunable multiple quantum well lasers, with distributed Bragg reflectors. The lasers operate at wavelengths near 1.3 μm. The outputs of the four lasers are combined by a passive waveguide structure. The device contains an optical amplifier section for the combined beams.

4. RECEIVERS

The functions of detection and amplification can be combined on a single chip as an optoelectronic integrated receiver. The material must be chosen to be compatible with both functionalities. Many demonstrations of integrated receivers have been performed. Figure 23-9 shows an example of a PIN photodetector and a field effect transistor (FET) amplifier integrated on an indium phosphide substrate. Such a device offers wide bandwidth and is potentially useful for applications such as lightwave communications.

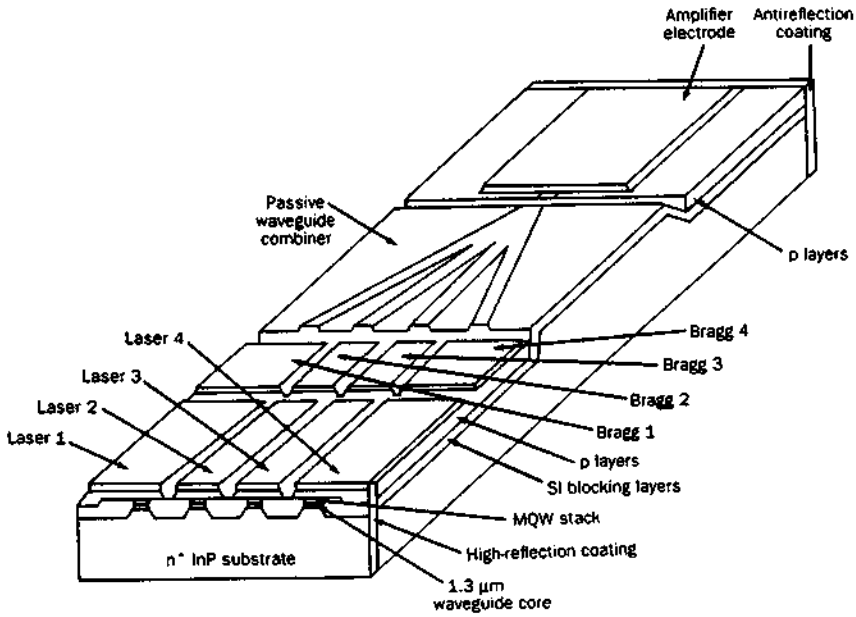


Figure 23-8 Four-laser wavelength division multiplexed source. (From T. L. Koch and U. Koren, *AT&T Tech. J.* 71, 63 (1992). Copyright © 1992 AT&T. All rights reserved. Reprinted with permission.)

D. Applications

The techniques of integrated optics can lead to optical systems configured as inexpensive, miniature integrated circuits for photonics applications. Although there have been many demonstrations of components suitable for integrated optic circuits and demonstrations of multiple components together on a single substrate, the application of integrated optics technology is still in a relatively early stage.

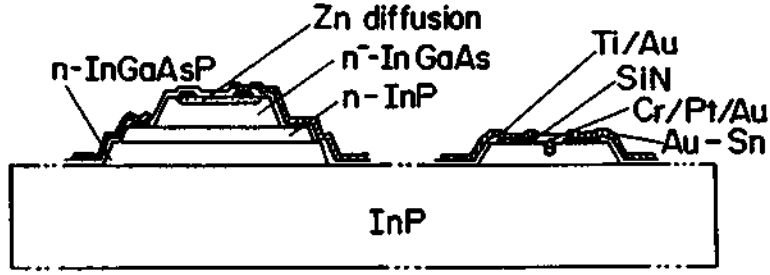


Figure 23-9 Optoelectronic receiver integrated on an indium phosphide substrate. The left portion of the structure comprises a PIN photodiode, and the right portion forms a field effect transistor amplifier. (From T. Horimatsu and M. Sasaki, *J. Lightwave Tech.* **LT-7**, 1612 (1989) © 1989 IEEE.)

The first large-scale applications of integrated optics are likely to occur in lightwave communications. There have been many demonstrations of integrated transmitters, receivers, and wavelength division multiplexers suitable for communication applications. The distributed Bragg laser sources widely used in lightwave communications could be considered as an integrated laser and resonator on a chip.

There are also likely near term applications of integrated optics for control of microwave devices. Such applications offer the advantage of isolation of the control signal from the microwave channels. Optical integrated circuits and microwave integrated circuits can be fabricated in the same materials. Examples of such optomicrowave monolithic integrated circuits have been demonstrated.

Integrated optics will likely be employed also in future generations of sensor systems, instrumentation, and data processing. Interconnection of very large scale integrated (VLSI) electronic circuitry has been demonstrated using integrated optics techniques. Another area that is being studied intensively is the use of optical interconnects to form a computer backplane, in order to reduce the number and size of electrical cables. The development of digital optical computing also will require integration of optical functions.

The field of integrated optics is in a state of rapid growth and development. A few examples of commercially available devices have appeared. Integrated optics will be used extensively in compact optical circuitry for future photonics applications.

References

[1] S. E. Miller, *Bell Syst. Tech. J.* **48**, 2059 (1969).
[2] D. Marcuse and E. A. J. Marcatili, *Bell Syst. Tech. J.* **50**, 43 (1971).
[3] F. K. Reinhart, D. F. Nelson, and J. McKenna, *Phys. Rev.* **177**, 1208 (1969).
[4] D. Hall, A. Yariv, and E. Garmire, *Opt. Commun.* **1**, 403 (1970).
[5] D. Mergerian *et al.*, *Appl. Opt.*, **19**, 3033 (1980).
[6] L. Thylen and L. Stensland, *IEEE J. Quantum Electron.* **QE-18**, 381 (1983).

Selected Additional References

R. G. Hunsperger, *Integrated Optics: Theory and Technology*, Springer-Verlag, Berlin, 1991.
L. D. Hutcheson, ed., *Integrated Optical Circuitry and Components*, Marcel Dekker, New York, 1987.
S. Martellucci, A. N. Chester, and M. Bertolotti, eds., *Advances in Integrated Optics*, Plenum, New York, 1994.
T. Tamir, *Guided-Wave Optoelectronics*, Springer-Verlag, New York, 1990.

Chapter 24 | Information-Related Applications of Lasers

Laser technology has become essential in the treatment of information and will become even more important in the future. Lasers have a wide variety of applications related to the processing, storage, and transmission of information. These applications involve a diversity of areas, like telecommunications, optical disks, optical computing, and display. In some cases, like lightwave telecommunications, laser-based techniques have revolutionized the way in which information is transmitted. In other cases, like optical computing, the most important laser applications remain in the future.

We shall consider a variety of these topics under the generic title of information-related applications. We consider specifically the areas of communication, data storage, optical data processing, and consumer products, such as point-of-sale scanners. The applications involving consumer products are especially noteworthy because they represent areas in which large numbers of persons employ or are affected by laser technology on a daily basis.

In what follows, we shall summarize a few selected examples that will illustrate the broad range of possibilities for lasers in information-related fields. The applications often employ some of the accessory equipment described earlier, particularly modulators, deflectors, and integrated optical assemblies.

A. Lightwave Communications

Communication using light as a carrier of information has an ancient history, extending back to signal fires and smoke signals, and perhaps reaching a previous peak in the nineteenth century with the use of the heliograph for military operations in the western United States. Communications using light faded in the late 1800s as electrical communications developed rapidly. But by the 1960s, the electromagnetic spectrum used for radio communication had become crowded. The development of laser sources at that time revived interest in optical communication technology.

The laser offered a coherent optical source with very wide bandwidth. Early ideas for laser-based communication systems generally involved free space transmission of a laser beam through the atmosphere, with the use of separate monolithic components including lasers, modulators, collimating optics, and so forth. A number of laser-based free space communication systems were demonstrated, based mostly on helium–neon lasers or gallium arsenide lasers. But because of problems with the turbulent and unclear atmosphere, free space laser–based communications remained largely experimental. It was soon recognized that the development of laser-based communication systems required use of a better transmission medium than the atmosphere.

The development of low-loss optical fibers, discussed in Chapter 22, provided the medium. Fiber optic elements had been known for many years and had been used to transmit light over short distances. Medical instrumentation useful for viewing internal portions of the human body had become available. But the relatively high values of loss associated with optical fibers had rendered them impractical for use in long distance communication. Until the late 1960s, the best optical fibers had losses in excess of 1000 dB/km. Over the period from about 1970 to 1980, the losses in optical fiber in the near infrared region were decreased by about three orders of magnitude and the ability to draw fibers in multikilometer lengths was developed. These developments made practical the use of optical fibers for telecommunication.

The term "lightwave communications" has been widely adopted to describe the use of optical energy to transmit information. This is consistent with other terminology, like microwave communications. Lightwave communications includes telecommunication systems based on fiber optics. The optical fiber is often called a lightguide.

A schematic diagram of a lightwave communications system is presented in Figure 24-1. It illustrates the relation between the various components. The information is generated as a digital pulse-code-modulated signal, and emitted as a series of pulses by the laser source in the transmitter. Typical loss levels for the optical fiber may be around 1 dB/km or less. In many systems, a total loss of 50 dB may be tolerated between the transmitter and receiver. Thus, a total distance up to 50 km is reasonable. When the loss becomes too high, a repeater is inserted, as illustrated. The repeater contains a detector and electronics to recover and amplify the signal, and a laser to retransmit it. Ultimately, perhaps after several repeater steps, the signal reaches the receiver that contains a detector and electronics and that recovers the digital signal. The figure also indicates the presence of a splice. Splices may be necessary for connection and field service. Splice technology for optical fibers has been developed, offering splicing with an insertion loss less than 1 dB.

It is instructive to consider the sources of signal degradation in an optical fiber. Figure 24-2 illustrates three leading causes. The figure shows the effects on the shape of an input pulse. The first cause is the attenuation, or loss, of the fiber, which may have a value of less than 1 dB/km. This reduces the amplitude of a pulse as it travels through the length of the fiber. When the loss is excessive, the presence of the pulse will no longer be detectable.

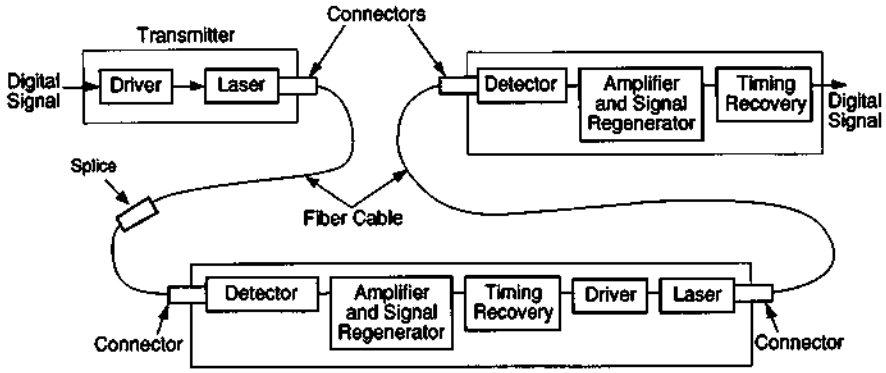


Figure 24-1 Lightwave communications link based on fiber optics.

The second cause is modal dispersion, which arises from the possibility of the optical energy taking paths of different length down the fiber. The pulse energy is not attenuated, but the pulse is spread out in time, so that after some distance it is no longer all within the proper time slot. As the spreading increases, it becomes impossible to detect the presence of a digital pulse in its proper time slot in a pulse-code-modulated communication system. As Chapter 22 described, this degradation may be minimized by use of single-mode fiber.

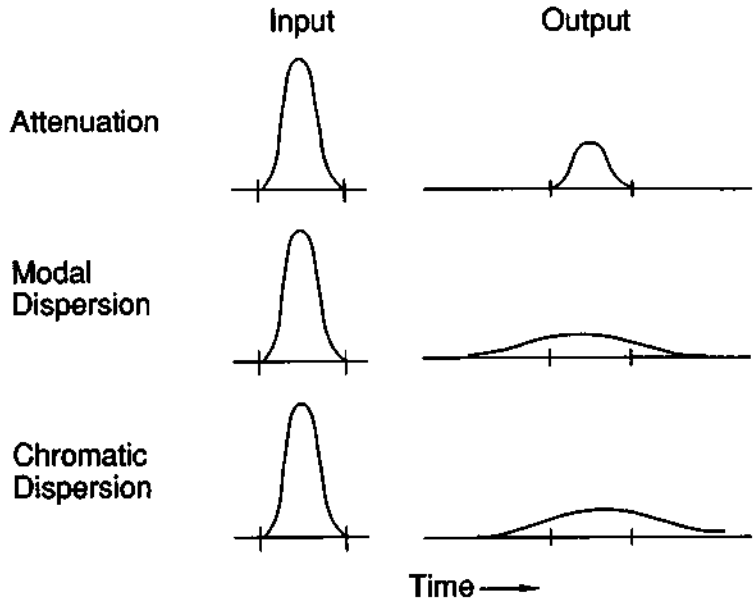


Figure 24-2 Causes of signal degradation in a digital fiber optic lightwave communications system.

The third cause is chromatic dispersion, which arises because of the finite spread of wavelengths within the pulse. Although the light may be highly monochromatic, there is always some finite spectral width. The shorter wavelengths arrive first after passage through the fiber, and the longer wavelengths arrive later. This also spreads the length of the pulse so that it does not arrive within the proper time slot.

Chromatic dispersion may be reduced by choice of laser wavelength. Figure 24-3 shows the variation of the chromatic dispersion with wavelength for the germania–silica-based fibers commonly used in fiber optic systems. The dispersion goes through zero near 1.3 μm. The bottom portion of the figure shows how the effect of chromatic dispersion broadens the pulse shape.

Figure 22-3 showed fiber loss as a function of photon energy over a fairly broad range. We now consider the loss as a function of wavelength in the region near 1 μm, the region important for lightwave communications. This is also the region where loss is near a minimum. Figure 24-4 shows the attenuation loss as a function of wavelength for modern single-mode optical fiber; the loss for multimode fiber is slightly higher. The peak near 1.4 μm is due to residual impurities. At shorter wavelengths, loss is dominated by Rayleigh scattering, which increases rapidly as one moves to wavelengths shorter than those indicated in the figure. At wavelengths longer than 1.6 μm, the loss is dominated by multiphonon infrared absorption, which increases rapidly as the wavelength gets longer. The 1 to 1.6 μm region has the lowest values of loss. Optical fiber loss is now very near the value set by intrinsic factors, like Rayleigh scattering, and is not likely to be reduced substantially below the values shown in Figure 24-4.

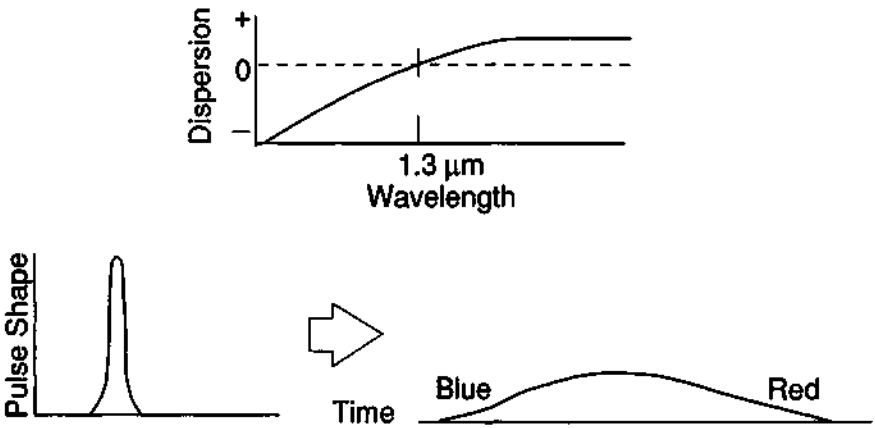


Figure 24-3 Top: Chromatic dispersion as a function of wavelength. Bottom: Effect of chromatic dispersion on pulse shape.

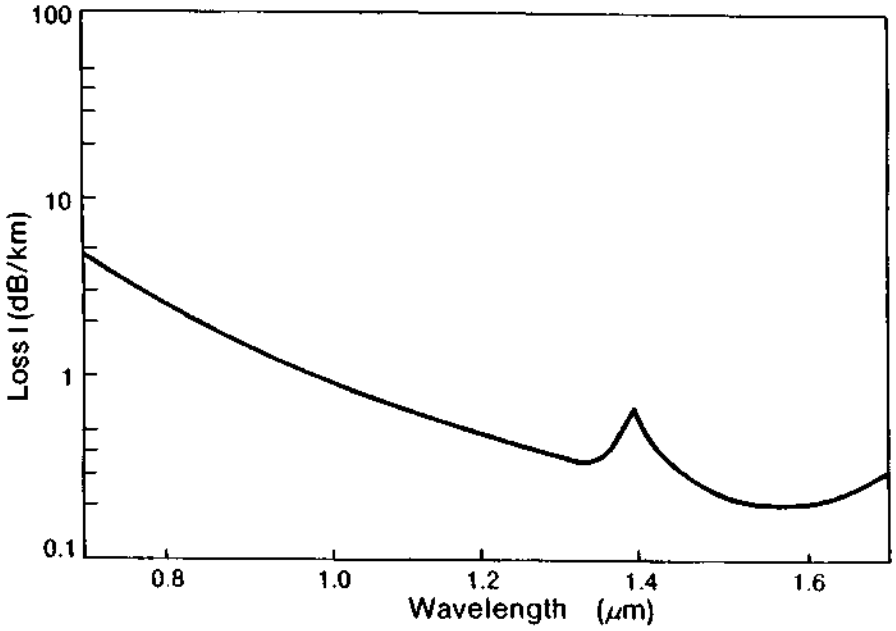


Figure 24-4 Optical fiber loss as a function of wavelength.

As optical fiber technology developed, the losses were reduced in different spectral regions at different times. The minimum loss is now at a wavelength near 1.55 μm. But historically, the losses were reduced to a low value first at a wavelength near 0.85 μm, a value that corresponded to the emission wavelength of aluminum gallium arsenide lasers. Thus, the first lightwave communication systems installed used these lasers and operated at a wavelength of 0.85 μm.

At a later time, as fiber development continued, the loss was reduced to a low value near 1.3 μm. Lightwave communication systems operating at this wavelength, using indium gallium arsenide phosphide lasers as sources, were developed. A large fraction of installed systems operate at this wavelength.

With further development, the minimum loss moved to a wavelength near 1.55 μm. Many modern systems now operate at the wavelength of 1.55 μm, using indium gallium arsenide phosphide lasers (of a composition different from the 1.3 μm sources). Most future systems are expected to use this wavelength also.

One important figure of merit for lightwave communications systems is the distance–bandwidth product, expressed in units of MHz-km. This figure of merit is the product of the data rate and the distance between repeaters (or the total system length, if there are no repeaters).

Systems that require relatively high performance, that is, high values of distance–bandwidth product, must employ components that provide high performance. There is a trade-off between the cost of the components and their performance. Systems

with relatively lower performance may use LEDs as sources, PIN photodiodes as detectors, and multimode fiber. Systems with high performance requirements will require laser sources, avalanche photodiode detectors, and single-mode fiber. An approximate dividing line between these requirements is a distance–bandwidth product of 100 MHz-km.

The development of diode-laser-based fiber optic lightwave communications systems has had a revolutionary impact on communication technology, particularly on long distance telephone communications. Lightwave communications grew explosively in the 1980s, going from experimental demonstrations at the beginning of the decade to routine use in telephone communications, with many millions of miles of installed fiber, by the end of the decade.

Fiber optic links provide significant advantages for telecommunications. These include low cost per channel, large system capacity, freedom from crosstalk and electrical interference, and light weight. In addition, because fiber cables are smaller than copper cables, more channels can be installed in existing crowded conduits.

Optical fiber lightwave communication systems provide data rates up to gigabits per second, at distances up to tens of kilometers between repeater stations. By the mid 1990s, there were about 60 million kilometers of fiber installed worldwide in lightwave communication links. Fiber optic systems are used extensively in intercity long-haul telephone links. They have been installed in transoceanic telephone cables. They have become widely used for interoffice trunk lines. In intracity links, they have replaced the more bulky copper cable in crowded conduits. Fiber optic systems have become the preferred medium for a wide variety of communication links.

We have so far emphasized long distance communications, characterized by large values of the distance–bandwidth product. There are many other possible applications for fiber optic systems, particularly for shorter distance communications. Figure 24-5 shows possible applications, characterized according to data rate and distance. The figure indicates data rates and distances appropriate for a number of applications.

The dashed line indicates a distance–bandwidth product of 100 MHz-km. Those applications lying to the upper right of the line are relatively high performance applications. They are basically telecommunication applications and may be regarded as the province of the common carriers. The long-haul telecommunication lines and interoffice trunk routes have become almost all fiber optic. Feeder lines, from central offices to remote terminals, have become partly fiber optic. Individual subscriber lines, the final distribution to homes, for example, have not yet been substantially replaced by fiber optics.

There are many lower-performance applications, lying to the lower left of the 100 MHz-km line. Those applications offer many opportunities yet to be exploited fully. For applications such as internal communications in ships, aircraft, and commercial buildings, fiber optics can replace copper wire and can provide substantial savings in size, weight, and cost. For applications like computer interconnection, fiber optics can offer very high data rates. Thus, in addition to the long distance telephone communication links, which have been very successful, numerous further applica-

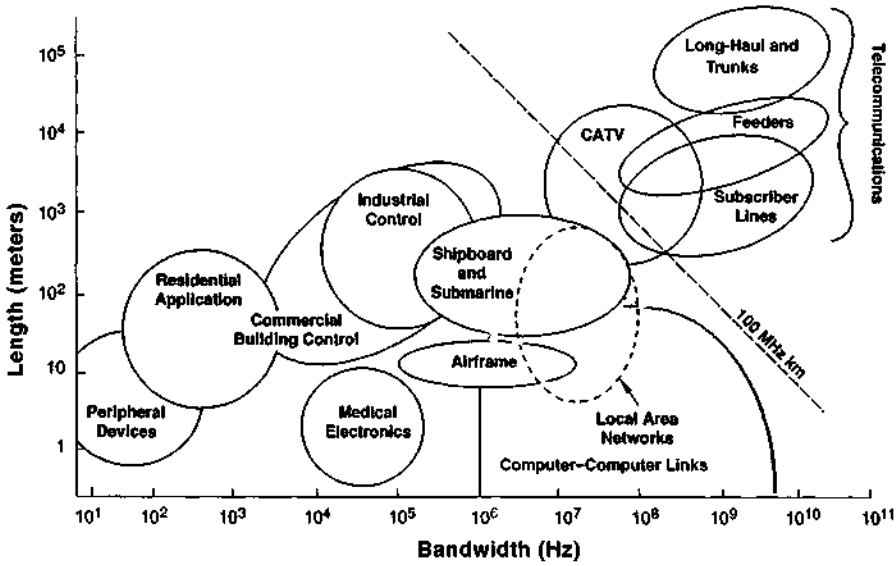


Figure 24-5 Applications for fiber optic communication systems, according to bandwidth (data rate) and distance. The applications lying to the upper right of the dashed line, which represents a distance–bandwidth product of 100 MHz-km, may be considered telecommunication applications. Those to the lower left of the line represent developing or potential lower-performance applications.

tions are possible. Substantial development work is under way to exploit these opportunities and to expand the role of fiber optic lightwave communications in new areas.

Although the exploitation of fiber optic lightwave communications has been impressive, there are still significant advances to be expected. One development that will aid high-capacity systems is the erbium-doped fiber amplifier (EDFA). These amplifiers consist of optical fiber containing ions of the rare earth element erbium. The fibers are pumped by a diode laser at a wavelength of 980 or 1480 nm. The requirements of the EDFA have provided a strong stimulus to the development of AlGaAs lasers operating at 980 nm, and such lasers have become commercially available.

The EDFA, when excited by the pump laser, provides optical gain at wavelengths around 1550 nm, the same wavelength region used by advanced lightwave communication systems. EDFAs can be used to amplify optical signals at that wavelength. They can provide optical amplification in a repeater without the need for optoelectronic detection and subsequent electronic processing and amplification. Thus, EDFAs can serve as the basis for all-optical communications, removing the requirement for electronics in the repeaters.

Although there have been some demonstrations of the use of EDFAs in operating lightwave communication systems, their application is still in an early stage. They do offer strong promise for the future.

Wavelength division multiplexing (WDM) is another area in which EDFAs offer advantages. In WDM, light from a number of laser sources, operating at slightly different wavelengths, is mixed into a single multicolor optical signal. The different beams are modulated independently so as to provide a number of different channels of information. The mixed signal is transmitted through an optical fiber, and the individual wavelengths are separated at the end by optical filters. Separate receivers then each receive the light from a single laser. The use of WDM allows multiple signals to be transmitted through the same fiber. This increases the number of channels that may be carried by a fiber.

Because EDFAs have a wide gain bandwidth, they may be used to amplify all the signals in a WDM system, simultaneously accommodating all the channels. The combination of EDFAs with high-density WDM should lead to more efficient use of fiber bandwidth and offer increased transmission capacity.

Another promising development of potential importance for optical fiber telecommunications involves soliton lasers. Solitons (or solitary waves) preserve their pulse shape and pulse duration over great distances. They are generated in media that exhibit both dispersion and optical nonlinearity, under conditions in which the pulse lengthening effects of dispersion are balanced by the pulse shortening effects of nonlinear optics.

Solitons have been generated in optical fiber lasers, which consist of fibers doped with rare earth elements. Soliton laser pulses have been produced with duration in the picosecond regime and with no change in pulse duration as they propagate. Transmission via solitons has been demonstrated at 20 gigabits/sec over distances of 13,000 km.

Soliton lasers offer promise for lightwave communications at higher bandwidth than is now possible. Although soliton laser technology is still developmental, it represents an extremely attractive alternative for very long distance communication.

In summary, the highly successful application of lasers in lightwave communication systems is still developing and expanding. It offers promise for higher performance in the future.

B. Optical Data Storage

Development of laser-based storage of data on various types of optical media began early in the 1960s, shortly after the invention of the laser. By the late 1960s, complete optical data storage systems had been developed, using helium–neon lasers and magnetooptical materials like manganese bismuth. These systems combined all the functions of recording, reading, erasing, and rewriting data bits. But the development of practical commercial systems lagged. The first products emerged in the late 1970s as read-only systems, using AlGaAs laser diodes, for audio compact discs. Now, a wide variety of optical memory formats are available.

Optical data storage takes the form of data bits (ones and zeros) recorded along circular tracks around the center of a rotating disk. The disks are fabricated from an

optically sensitive recording material. The data are recorded in the form of small spots, with dimensions of the order of the wavelength of light. The spots can change the reflection or polarization of light. They may be written by a pulsed laser, focused to a diffraction-limited spot. The laser beam heats the material and changes its properties. A site with changed properties could represent a one bit, and a site with unchanged properties a zero bit. Data are retrieved as the disk rotates and the stream of bits passes under the read head. The reading is performed with a lower-power laser. The change in reflection or polarization is detected and used to reconstruct the sequence of ones and zeros.

Optical memories are bit serial devices, not random access devices. They read data as a stream of bits passing the read head. In general, to reach a given bit, one must wait approximately half the period of rotation of the disk, on the average. But once the beginning of a record is reached, the data in it can be retrieved at a high rate.

Optical memories have taken three main forms.

Read-only memories In these devices, the data is represented by indentations in a plastic disk. The indentations are embossed in the disk during an injection molding process that forms the disk. Such disks have been used primarily for entertainment. The popular audio compact disc is of this type, and video compact discs are becoming available for movies. These devices will be described more fully in a later section on consumer products.

Write-once, read-mostly (WORM) memories These devices use a disk coated with a material that melts at a low temperature. Tellurium alloys have been used. The writing laser burns a pit in the coating to record a bit. The presence of the pit changes the reflectivity at that location. The change in reflectivity is sensed by the reading laser. Once recorded, these disks cannot be erased. That fact is not always a disadvantage, because it provides security for the data. Such devices have become widely used to record data for banking and for the health industry. They are also used for archival storage of large amounts of data and as digital map systems in military aircraft.

Figure 24-6 shows one configuration for a write-once, read-mostly memory. The reflecting surface is originally coated with a layer with relatively low reflection. The laser diode may be modulated to emit a relatively high level of power when data are to be recorded. The laser produces pits in the material on the reflecting surface. When this material is removed, light is reflected well by the reflecting surface. The reflection of light from the surface changes when light is reflected from a pit or from an area of the original material. This change in reflected light identifies the presence of a one or a zero at a particular position. Because the data are stored by vaporizing pits in a material, they cannot be altered after they have been written.

When the data are to be read, the laser power is reduced. The quarter-wave plate rotates the plane of polarization of the light by 90° after a double passage through it. Light reflected from the surface is then sent by the polarizing beam splitter to the photodiode detector. The detector is shown as a quadrant device. The optics

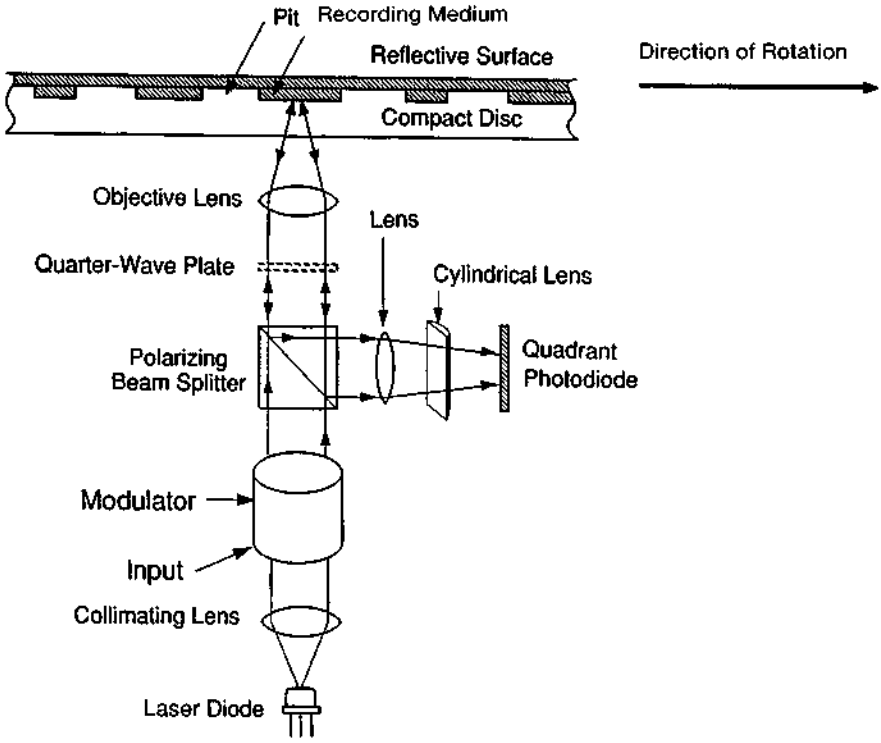


Figure 24-6 A possible configuration for a write-once, read-mostly optical memory.

purposely have some astigmatism, so as to send the beam to different positions on the quadrant detector as the beam focus changes along the track of bits. This aids the function of controlling the beam focus accurately.

Rewritable memories In these devices, data may be recorded, erased, and rewritten many times. These disks have a coating of a magnetooptical material, frequently a matrix of transition metals (like iron and cobalt) and rare earth elements (like gadolinium and terbium). The writing laser heats a small spot on the disk and changes its state of magnetization. This in turn changes the polarization of the beam from the reading laser. The change in polarization is transformed to a change in light intensity by a polarizer, and this change is sensed by a detector. Figure 24-7 shows a possible configuration for a magnetooptic memory device. Rewritable optical disks offer very high areal density of data and are useful in storage-intensive applications, like desktop publishing.

Let us describe the operation of rewritable memories in more detail. The magnetooptic material is initially magnetized to saturation in one direction. The focused laser light produces localized heating, raising the temperature in a small area above the Curie point, the temperature at which the magnetism of the material is destroyed.

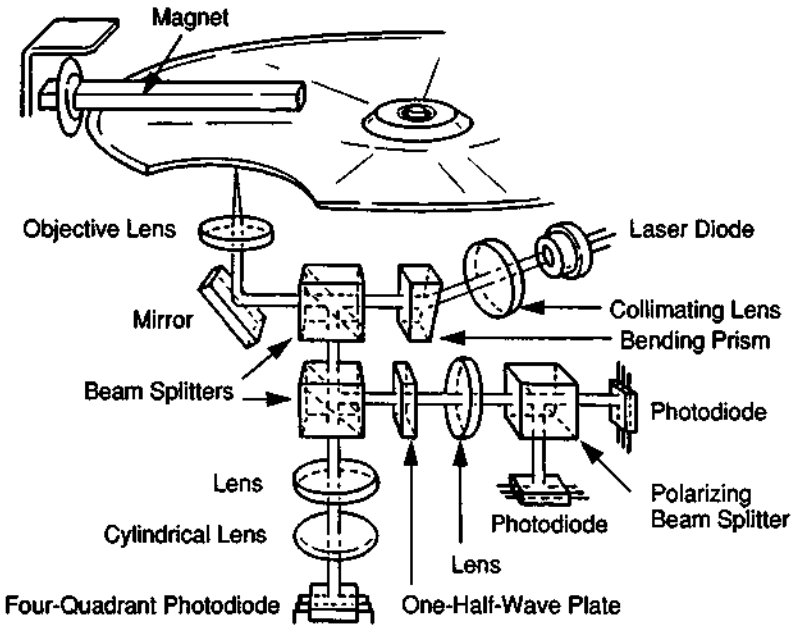


Figure 24-7 A magnetooptic memory device. (From *Laser Diodes*, Sharp Corp., 1992).

For typical magnetooptic materials, this is around 200°C. When the pulse of light ends, the material cools and becomes remagnetized in the direction of the magnetic field supplied by a bias magnet. Remagnetization in one direction can represent storage of a "one" bit of information; remagnetization in the opposite direction represents a "zero."

Readout of the information uses magnetooptic effects arising from the interaction of polarized light with magnetic materials. The polarized light reflected from the material has its direction of polarization rotated slightly, perhaps around 0.4° in a typical situation. Although this is a small value, it is enough to produce a change in the intensity of the light transmitted through a polarizer oriented so as to extinguish the light. The contrast can be high because the signal is detected against a near-zero background. The direction of the rotation depends on the direction of the magnetization of the material. The magnetooptic effect provides a method of sensing the direction of magnetization in a small area on the material and, thus, can determine whether the information in that area represents a one or a zero.

Most optical data storage devices to date have used AlGaAs semiconductor diode lasers operating at a wavelength near 780 nm. These diode lasers must be of high quality, with good wavefront uniformity, long lifetime, and low noise. For reading, an output of a few milliwatts is usually sufficient, but for writing and erasing, a few tens of milliwatts are required.

Most of the lasers made worldwide are for use in optical disks. Tens of millions of diode lasers are produced annually for use in compact disc players. This is by far

the largest use of lasers in the world. In addition, hundreds of thousands of lasers are used annually in rewritable magnetooptic devices.

One of the attractive features of optical data storage is the very high areal density of data on the disks. The areal density of the data storage increases as the inverse square of the laser wavelength, because the data are stored in diffraction-limited spots. Thus, there is a strong desire to move to laser sources with shorter wavelength. Some systems are becoming available using 680 nm lasers. This change increases the density of data by about 30 percent. The availability of blue or violet semiconductor lasers would further increase the density greatly. This fact strongly motivates the development of short-wavelength semiconductor lasers. We may anticipate the use of such lasers in optical data storage devices within a few years.

Figure 24-8 presents a projection for the increase in areal density available in rewritable optical storage devices as a function of time. The density may be expected to increase from hundreds of megabits per square inch in the mid 1990s to near 10 gigabits per square inch in the early twenty-first century.

Optical disk memories compete with magnetic disks and with semiconductor memories for data storage applications. Each type of memory device has advantages for specific applications, so that all the types of memory will continue to be used. Semiconductor memories have very fast random access to data, which disk memories lack. Magnetic disks offer somewhat higher data rate and faster access time than optical disks, at least in the current status of development.

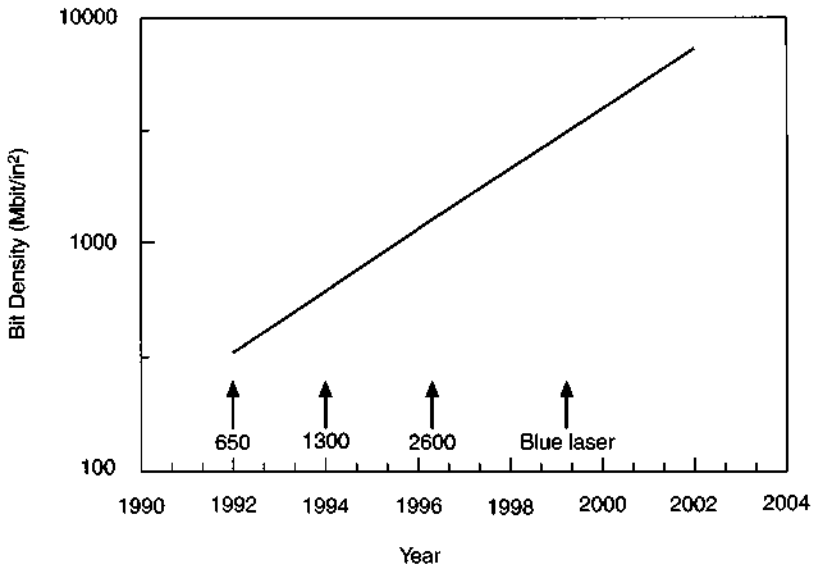


Figure 24-8 Projected evolution of bit density in optical data storage devices. The arrows with numbers represent the introduction of 5.25-in.-diameter disks with the indicated total number of megabytes of data. The arrow marked Blue laser indicates the anticipated introduction of blue lasers as sources.

The packing density of optical storage is very high, as Figure 24-8 has indicated. The packing density is higher by about a factor of ten than for magnetic disks. For very large memories, the cost of optical storage is substantially lower than for semiconductor devices. The advantages of optical memories favor applications involving highly intensive storage needs.

Optical disk memories also offer the advantage of easy removability. The read and write heads in an optical disk system are separated from the recording medium by a distance of around 1 mm. This is in contrast to the magnetic disk, in which the head and medium are almost in contact. This factor eliminates head crashes and allows for easy removal and replacement of the disk.

Optical disks thus offer a combination of advantages that favor their use in situations where either removability or very high volume storage are important. These applications include items like replacement of floppy disks or very large "jukeboxes" containing multiple disks. As optical data storage technology continues to evolve and mature, the performance parameters will continue to improve, and applications will continue to expand.

The optical data storage technology discussed so far has all been bit oriented, that is, a small area on the recording medium represents one bit of information. There is another possibility, holographic optical memory, in which the information is stored as holograms, each of which represents a large number of bits. Holographic optical memory systems have been under development since the 1960s, almost as long as the bit-oriented systems. They have not yet progressed to the point of broad commercial use, as optical disks have, but in the early 1990s there has been a resurgence of interest in them. For some applications, they offer significant advantages compared with bit-oriented memories.

The organization of a holographic memory is different from that of a bit-oriented memory. In a bit-oriented memory, information storage and readout occur one bit at a time. A holographic memory stores and reads out a large number of bits simultaneously. One configuration for a holographic memory is shown in Figure 24-9. The apparatus in the figure contains a holographic recording arrangement, with a reference beam and a signal beam. The information is stored as a hologram in the holographic storage medium. The object from which the hologram is formed is a two-dimensional array of bits. This array is constructed by a device called a page composer, which is a spatial light modulator (see Chapter 5). The page composer is an array of light valves, some of which are open and some closed. The opening and closing may be activated by electrical signals, by light, or by both. The open valves can represent ones, the closed valves zeros. The pattern of the light valves is set to represent the desired array of ones and zeros. This array of ones and zeros is stored at one time in the holographic storage medium. Because the hologram contains many bits, one has simultaneous storage of a large number of bits.

The laser beam is divided by the beamsplitter into a reference beam and a signal beam. The signal beam is directed to the page composer, where an electronically composed data pattern is set up. This pattern is imposed on the signal beam. When the signal beam combines with the reference beam to form a hologram in the recording medium, the hologram represents the entire array of bits. The hologram is

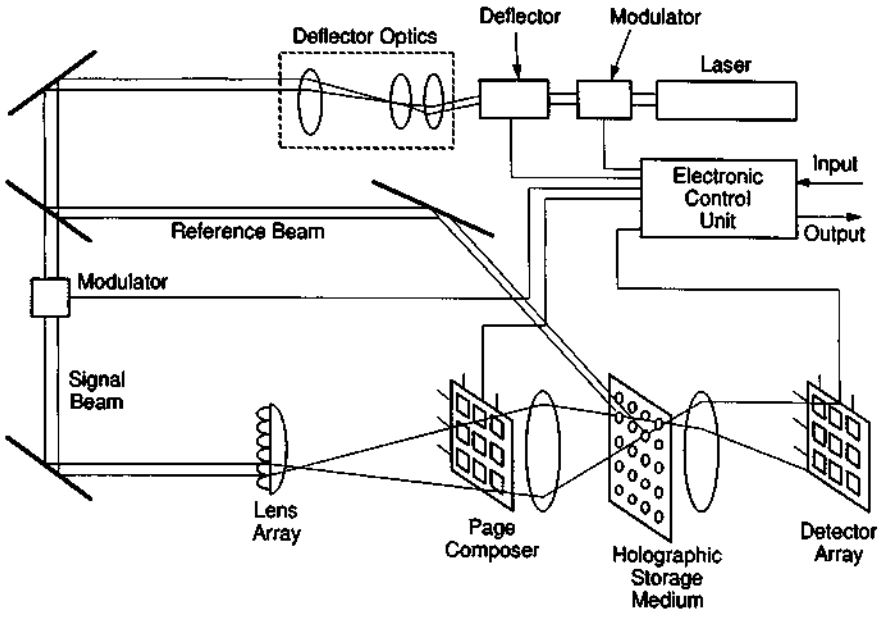


Figure 24-9 A holographic memory system.

formed on one particular small area of the storage medium, an area selected by the light beam deflector.

To store a different hologram at another location, the deflector moves the beam to that location. Movement of the beam from one lens to another in the lens array changes the position of the hologram on the storage medium. At the same time, the reference beam tracks the signal beam so that both beams reach the same spot. If one changes the angle between the two beams, it is also possible to store multiple holograms in the same area on the memory material.

Readout of data occurs when the hologram is addressed only by the reference beam. The modulator may change the light intensity between recording and reading. The deflector directs the beam to the hologram that is to be read out. An image of the array of bits is produced and focused by a lens onto an array of photodetectors, which has the same number of elements as the page composer. Each bit from the page composer is incident on one photodetector in the array. The data thus are converted back to an electrical signal. If a particular area on the page composer represented a one, there will be light incident on the photoconductor in the corresponding position is the photodetector array. Thus, the array of bits can be reconstructed and converted to an electrical signal in parallel, with all the bits on the page being recovered at the same time. This feature allows the readout rate to be very high.

Storage of data in holographic form offers several advantages compared with a bit-oriented memory. The information is distributed in a holographic fringe pattern that covers the entire hologram. Thus, the storage is not sensitive to the presence of small imperfections, which could cause the loss of bits in a bit-oriented memory.

A second advantage is that the holographic recording is relatively insensitive to the exact position of the beams on the recording medium. This is in contrast to the bit-oriented memory, in which the light must be positioned very exactly.

Another advantage is the very high readout rate that results from the parallel nature of the system. For a 10^9-bit system (10^4 pages of 10^5 bits each), one requires only 10^4 locations on the storage medium. This figure lies within the capability of high-speed inertialess light beam deflectors. Thus, addressing can be done entirely by nonmechanical light beam deflectors, which have random access times less than 10 μsec. The photodetector array would require about 316 by 316 detectors. The system would have no moving parts. Because 10^5 bits are stored (or read out) at the same time, one could have data rates in excess of 10^{10} bits/sec.

Despite these advantages, the development of holographic optical memory substantially lags behind that of bit-oriented memories. One problem has been the availability of suitable recording materials, which can be rapidly erased and rewritten many times. A second problem has been the page composer, which represents a large array of light valves with stringent requirements. There does exist serious interest in holographic memory systems, and prototype systems are under development by a number of organizations. A substantial amount of research is under way to relieve the problems. It appears likely that practical holographic optical data storage systems will become available in the relatively near future. As of early 1996, the first commercial products are becoming available. One model containing 1000 pages, each containing an array of 640 by 480 bits, and capable of being read at rates in excess of 300 megabits/sec is being marketed.

C. Optical Data Processing

Optical methods can be used in a variety of ways for processing of data, in other words, for optical computing. There have been many examples of techniques that process information in the form of distributions of light intensity in a parallel fashion for performance of mathematical operations. We term these applications analog optical computing. Other approaches involve use of optical components such as gates, switches, and flip-flops in an architecture that processes data in digital format and that operates similarly to an electronic computer. We term these approaches digital optical computing.

Analog optical computing has been developed over a period of years. The techniques are well established, and the technology has applications in areas like pattern recognition and image processing. Digital optical computing is still in an early stage of development.

1. ANALOG OPTICAL COMPUTING

Analog optical processing is highly parallel in nature and often involves processing data that can be represented in pictorial format. Various optical techniques have been developed to carry out certain types of mathematical operations. These operations include performance of Fourier transformations and convolutions.

Many processing operations exploit the fact that a lens produces a Fourier transform of the light distribution present in its front focal plane. The light distribution can be defined by a photographic transparency placed in the front focal plane. The transform appears in the back focal plane of the lens. If an amplitude mask is placed in the back focal plane, then the convolution of the Fourier transform of the mask with the amplitude distribution of the original transparency appears in the back focal plane of a second lens. If the transparency and the mask have amplitude distributions that represent two-dimensional mathematical functions, these optical techniques produce the mathematical operations of Fourier transformation and convolution on the input functions. The output function is represented by a two-dimensional distribution of light, which may be photographed or sampled by photodetector arrays.

A common arrangement is shown in Figure 24-10. Two lenses of equal focal length F are separated by $2F$. The input plane is the front focal plane of the first lens. The data to be processed can be inserted as a photographic transparency in this plane. The rear focal plane of the first lens (also the front focal plane of the second lens) is called the transform plane. The rear focal plane of the second lens is called the image plane; the processed data is recovered there. One may insert a viewing screen, photographic film, or an array of photodetectors in this plane.

A collimated beam of light is required. In practice, the light source will almost always be a laser. It is possible to carry out the operations described here with nonlaser sources, but usually the high radiance of the laser makes it the most practical source.

Many analog optical data processing applications rely on the ability of a lens to form the Fourier transform of a distribution of light. In the figure, the distribution of light in the input plane is given by $g(x_1,y_1)$, where x_1 and y_1 are the spatial coordinates in the input plane. The light distribution may be produced by a photographic transparency inserted in this plane. Then, the light distribution G in the transform plane is the two-dimensional Fourier transform of g:

$$G(p, q) = \iint_{-\infty}^{\infty} g(x_1, y_1) \exp[i(px_1 + qy_1)]\, dx_1\, dy_1 \tag{24.1}$$

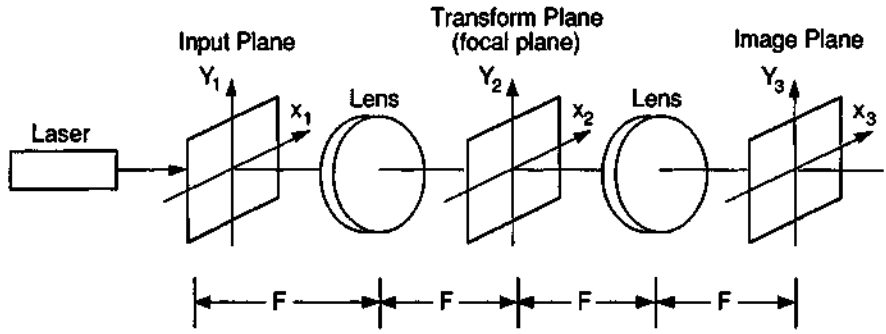


Figure 24-10 Analog optical data processing.

where p and q are spatial frequency variables related to the spatial coordinates x_2 and y_2 in the transform plane by

$$p = \frac{2\pi x_2}{\lambda F}, \qquad q = \frac{2\pi y_2}{\lambda F} \tag{24.2}$$

with λ the wavelength of the light. These variables have dimensions of radians per unit length.

In Equation (24.1), a multiplicative constant, which does not affect the physical interpretation, is neglected. This Fourier transform relationship is the basis for many filtering operations that lie at the heart of optical data processing.

The second lens in the system will form a Fourier transform of the light distribution G in the transform plane, thus re-forming the original function. The light distribution $g(x_3, y_3)$ in the image plane will be

$$g(x_3, y_3) = \iint_{-\infty}^{\infty} G(p, q) \exp[i(px_3 + qy_3)]\, dp\, dq = g(-x_1, -y_1) \tag{24.3}$$

This equation is fundamental to analog optical data processing. The operation is not the inverse Fourier transform, because the sign in the exponential is the same as in Equation (24.1). Still, the successive application of two Fourier transforms reforms the original light distribution, but with the coordinate system reversed. The inversion of the image is expected for the optical system shown in Figure 24-10.

The transformation has several interesting properties. The light that corresponds to details with fine structure (high spatial frequency) in the original light distribution g appears relatively far from the origin in the transform plane. Details corresponding to a slowly varying light distribution will be focused closer to the origin. This leads to the possibility of removal of unwanted features by physically putting apertures in the transform plane.

The optical processing operation involves filtering part of the light distribution in the transform plane. The simplest filters are binary amplitude filters, having transmission of either zero or unity; that is, they are simply stops for part of the light distribution G. Such filters are easy to fabricate. Other filters may be more complicated, and they may operate on both the amplitude and phase of the light distribution G.

The information to be processed is assumed to be in the form of a photographic transparency. It could, for example, represent data from weather satellites, on which screening or pattern recognition operations were to be performed. It could represent aerial photographs for which processing could involve the recognition of features or removal of uninteresting features to make later photointerpretation easier.

If no processing is performed by apertures in the focal plane, the image will be the same as the original light distribution, but inverted. If processing is performed by removing some portion of the light, then the image will have some features removed. The features that are removed are those that correspond to the light that is blocked.

If, for example, one wanted to remove wrinkles from a photograph of a human face, one would insert a low-pass filter in the transform plane. This filter will pass

light near the origin but remove light farther from the origin. The wrinkles correspond to a fine structure, which would be focused relatively far from the origin. This light would be blocked. When the image is re-formed, the low-frequency portions of the photograph will be present, but the wrinkles would be absent.

If an opaque stop is placed at the origin of the transform plane, then light corresponding to broad background is eliminated. This procedure will accentuate details, making fine structure stand out in the presence of slowly varying background.

Let us consider another example. Suppose we have a scene with an overlapping set of parallel horizontal lines. This is illustrated in Figure 24-11. The light passing through the transparency will have a distribution g that contains the desired scene and the interfering set of lines. When the distribution of light is Fourier transformed by the lens, the array of horizontal lines will give rise to a diffraction pattern consisting of a series of spots in the vertical direction, as shown in the lower left in the figure. If one inserts a mask of the form shown in the lower right, the vertically oriented spots are blocked. The second lens then re-forms the image of the original scene with the horizontal lines removed. This application has in fact been used to remove interfering raster lines from photographs sent to earth by spacecraft.

This example is meant to show the utility of optical data processing in dealing with information presented in pictorial format. There are many more sophisticated types of filtering operations that can be performed in the transform plane. These in-

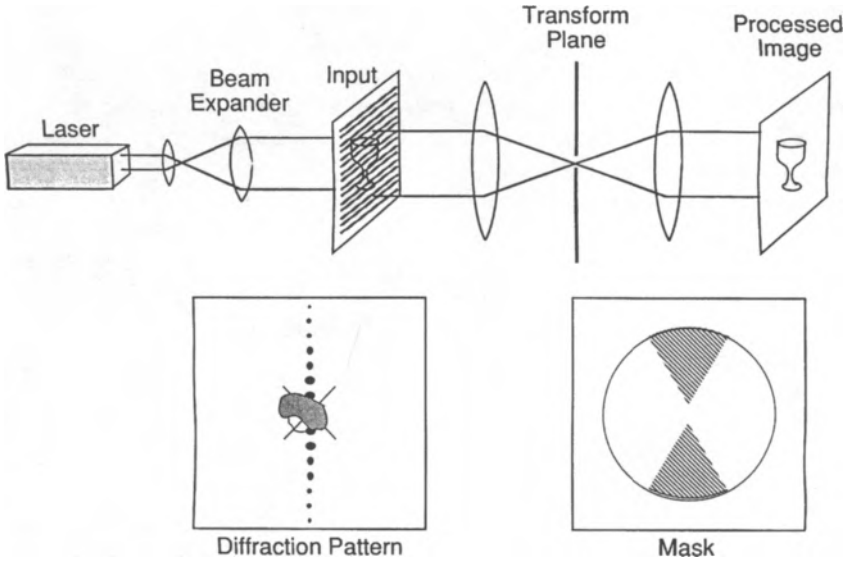


Figure 24-11 Use of optical data processing system for removal of unwanted horizontal lines in a scene. The lower left inset shows the distribution of light in the transform plane. The lower right inset shows the form of the mask required to filter light corresponding to the horizontal lines.

clude filters that operate on both the amplitude and the phase of the light in the transform plane. It is beyond the scope of this work to describe all the variations. We will mention one especially significant possibility, the matched complex spatial filter. This filter allows automatic inspection of photographic data and retrieval of specific features in the input scene. Such processing can be used for searching of aerial photographs for objects of a specific shape. It also can be used for optical character recognition. If a matched filter for a given letter, for example, *e*, is placed in the transform plane and the input consists of alphanumeric characters, the output will be a number of spots of light, one at each position where the letter *e* appeared in the input.

One method for fabrication of a matched complex spatial filter is shown in Figure 24-12. The method essentially produces a hologram of the object or character that one desires to recognize. Assume that one desires a matched complex spatial filter for the letter *e*. The input would be a transparency of the letter *e*. The hologram is formed on the photographic plate, which is then developed and which yields a matched complex spatial filter for the letter *e*. When this filter is placed in the transform plane and a page of printed data is used as the input, spots of light corresponding to the location of each occurrence of the letter *e* appear in the image plane. This is shown in Figure 24-13, in which part (a) shows the input and part (b) shows the occurrence of spots of light at positions corresponding to the locations of the letter *e* in the original material.

In this example, the image is not faithfully retained, that is, the output is not in the form of the letter. This is because the phase of the light is not faithfully preserved in the transform plane. The technique may still be useful for detecting the presence and position of a signal without having to preserve its exact shape. The advantage of matched filtering is that the result gives a sharp peak of light, even if the original signal has no peaks. Thus, an array of photodetectors with threshold logic can detect the spots of light. The broader, more diffuse distributions of light arising from partial correlation with other members of the character set are rejected by the threshold logic.

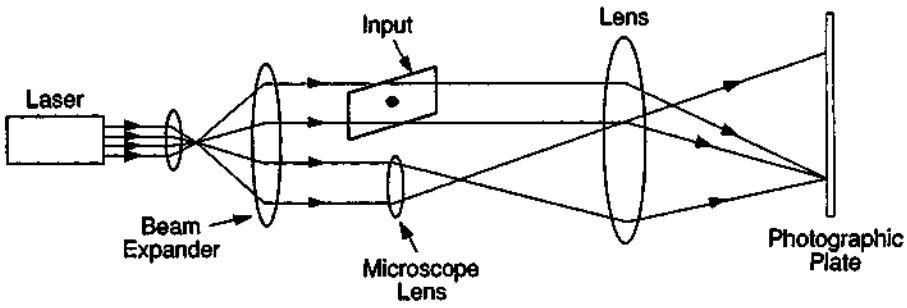


Figure 24-12 Formation of matched complex spatial filter.

Figure 24-13 Identification of the letter *e*. The input is shown at top. Identification of the locations of the letter *e* are indicated at bottom by spots of light. (From S. Lowenthal and Y. Belvaux, *Opt. Acta* **14**, 245 (1967).)

Although the fundamentals of optical information processing were developed in the pre-laser era and a number of techniques were demonstrated with incoherent sources, the creation of matched spatial filters requires a high degree of coherence and, thus, requires laser sources.

Despite early enthusiasm, use of this technology for optical character recognition has not been widely adopted. Optical character recognition is very sensitive to the exact size and orientation of the letters. The system requires many filters in order to identify all the possible characters. It is also sensitive to changes in type font; a character in one font may not be detected well by a filter derived from the same character in a different font. These drawbacks have kept this form of optical character recognition from widespread application.

Some limited applications of pattern recognition of practical significance in industry have developed. One application is the automatic recognition of defects in patterns. This has been applied for rapid inspection of large numbers of similar objects, like repeated patterns on a photolithographic mask. Suck masks contain large numbers of identical small patterns, making complete inspection by conventional techniques difficult. Because the mask is an array of the same pattern repeated at the same spacing, it acts like a diffraction grating. In the transform plane, there will be an array of spots spaced at the periodic frequency of the pattern. The amplitude filter in the transform plane can consist of an array of opaque dots, at positions corresponding to the spots characteristic of the periodic structure. The dots will block all light that is repeated at the period corresponding to the mask spacing. But random errors that are not repeated will cause light to miss the dots and reach the image plane. This technique has been used for rapid detection of random errors in photomasks.

Methods for optically performing mathematical operations other than those we have discussed have also been developed. For example, methods for performing the basic operations of addition, subtraction, multiplication, and division, as well as calculation of convolutions, correlations and power spectra, differentiation and integration, and Hilbert transformation have been devised. These methods offer promise for optical computation related to two-dimensional functions represented by distributions of light intensity. Data obtained in a two-dimensional form (like photographs) could potentially be processed in a very natural way. For data that is available in an appropriate form, the parallel nature of the processing offers advantages when compared with the sequential nature of electronic processing. When large amounts of data are available, analog optical processing could be much faster than conversion to a digital form with subsequent electronic processing.

Applications that have been demonstrated include deblurring of photographs, processing of radar signals, matrix multiplication, and machine vision. Although some practical applications for optical data processing have appeared, especially applications involving Fourier transforms, widespread application of this type of computing still lies in the future. The techniques offer substantial promise, especially for parallel processing of large amounts of data already available in an optical format, but more developments remain to be done, even though the basic principles have been established for a number of years.

2. DIGITAL OPTICAL COMPUTING

In addition to the analog computing applications just discussed, there is substantial research interest in the development of digital optical computing. Digital optical computers would offer advantages over electronic computing. Optical computing could in principle be very fast. Electrons must travel along conductors, whereas photons do not. Any transmission of an electronic signal along a wire involves charging a capacitance, which depends on the length of the wire. Optical signals do not require the charging of a capacitance and, thus, can have larger bandwidth. Because photons do not need to be guided by conductors, it is possible to establish a level of interconnection that would not be possible with electronic circuits. Optical beams can pass through each other without interference, which is not true of electrons. Because photons have no electrical charge, they are immune to electromagnetic interference. The combination of global interconnection, high level of parallelism in the processing, and potentially very high speed could lead to very great computational power.

A digital optical computer would require optical components to perform the functions of the switches, gates, and memory elements that control the flow of electrons in an electronic computer. The functions of the gates and switches may be provided by optical modulators, which have taken many forms. A 1×1 switch is simply an on–off switch that can connect two lines. A 1×2 switch can connect one line to either of two lines. A 2×2 switch connects two lines to two lines. It has two states, with either input 1 connected to output 1 and input 2 connected to output 2, the so-called bar state, or input 1 connected to output 2 and input 2 connected to output 1, the so-called cross state. Hence, it is called a crossbar switch. A crossbar switch can be extended to an $n \times n$ configuration, in which any of n input lines can be connected to any of n output lines at will without interference.

Optical switches may be formed from optical modulators (light valves) of many types, such as electrooptic modulators, acoustooptic modulators, magnetooptic modulators, and liquid crystals. An array of such modulators can be used to form a spatial light modulator, as was discussed in Chapter 5.

An array of modulators can also be used to form a crossbar switch. For example, an $n \times n$ crossbar switch could be formed by an array of $n \times n$ light valves. A vertical linear array of n laser diodes would serve as the light source. A cylindrical lens spreads the light from the diodes horizontally, so that each diode illuminates one row on the $n \times n$ array. Then, a set of n cylindrical lenses, oriented perpendicular to the first lens, is used, one lens per row of the modulator array, to focus the light transmitted by the array onto a linear array of n photodetectors, oriented horizontally. This configuration can connect light from any laser diode to any detector without interference.

Both electrically and optically controlled switches have been developed. For an all-optical computer, one would employ optically controlled devices. One also desires that the switches be small, high-speed devices capable of being fabricated in large arrays, and that they require very low switching energy. The desired switching functions have been demonstrated using several different technologies. But the fabrication of large, high-density arrays of optical gates is still developing.

The function of the memory elements (flip-flops) in an optical computer can be performed by devices exhibiting optical bistability. Optical bistability refers to optical devices with characteristics that can take on either of two distinct states, depending on the previous history of the device. For specificity, we will consider the transmission characteristics of the device, although other properties may also by employed. Figure 24-14 shows how the light intensity transmitted by an optically bistable device can vary as a function of incident light intensity. In the left portion of the figure, the transmitted intensity increases with incident intensity until some threshold value is reached. The transmitted intensity then switches up to a higher value. As the incident intensity is later reduced, the transmitted intensity remains in its high state until another, lower threshold value is reached. At that point, the transmission switches down again. There are two stable states for the transmission, a high value and a low value. Switching between these two values may be accomplished by brief changes in the incident light intensity. The form of the curve is similar to a magnetic hysteresis loop.

The right portion of the figure illustrates that the bistability may work in the reverse sense of what was shown in the left portion. The transmission switches down as the switching threshold is reached. It remains in its low state as incident intensity is decreased, until the lower threshold is reached. At that point, the transmission switches back up again.

Many types of optically bistable devices have been demonstrated. As an illustrative example of a bistable device, consider a Fabry–Perot interferometer with a medium between the mirrors the index of refraction of which depends on light intensity. At low values of light intensity, the index has value n_0. For light incident normal to the mirrors, the interferometer has transmission maxima at values of wavelength λ that satisfy

$$N\lambda = 2n_0 D \tag{24.4}$$

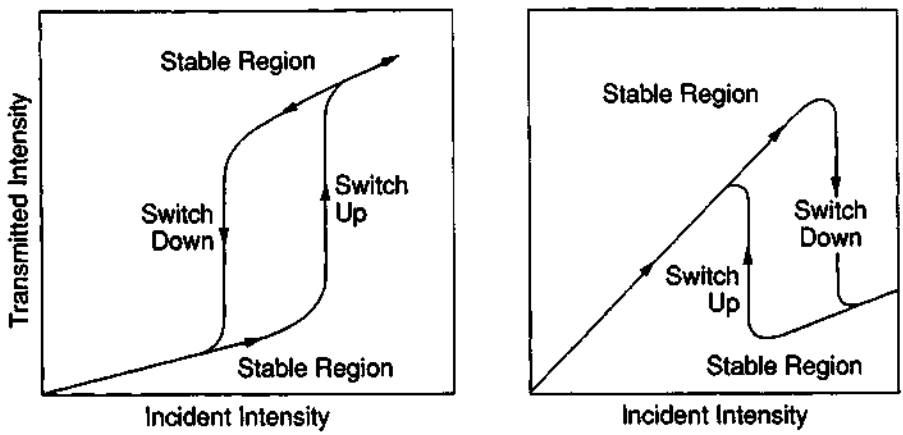


Figure 24-14 Optical bistability.

where N is a large integer and D is the distance between the mirrors. Consider a particular combination of values of N and λ for which this equation holds. For low values of incident light intensity at this wavelength, the transmission of the device will be high. But as the incident intensity increases, the value of the index of refraction will also change, and the equation will no longer be satisfied. Then, the transmission will drop to a low value. Later, as the incident intensity is reduced, the index of refraction will return to its original value n_0, the equation will again be satisfied, and the transmission will return to its high value.

Another important type of optically bistable device is the so-called self-electrooptic-effect device (SEED). The SEED is fabricated from a multiple quantum well structure, consisting of alternating thin layers of different types of semiconductor materials, like GaAs and AlGaAs. An external electric field is applied to the device. The absorption of the device is a nonlinear function of the applied field in the layers. But the field changes with light intensity, because absorption of light creates electrical charges that change the electrical conductivity of the material. Thus, the optical absorption changes as the incident light intensity changes. The result is optical bistability.

SEEDs have been notable because they can be switched very rapidly, in the nanosecond regime, and require very small switching energy, as low as femtojoules (10^{-15} J, fJ). Thus, they appear desirable for optical computing applications.

Optical bistability has been demonstrated using many different materials in a variety of device configurations. The availability of optically bistable devices allows for optically controlled components useful for optical computing, such as optically activated switches, gates, and memory devices.

Thus, the components needed for digital optical computing have been developed and demonstrated, although not necessarily in the large arrays required for a full-scale optical computer. Many demonstrations of all-optical processing have been performed. We shall present a single example, out of many possible, to illustrate the status of optical computer development and the architectural concepts.

The example is a digital electrooptic computer testbed called DOC II [1]. This device has been used to perform general purpose computing. It uses 64 individually controlled laser diodes as sources. The beams from these lasers are directed by a series of anamorphic lenses to a spatial light modulator, which is fabricated from gallium phosphide acoustooptic devices. The spatial light modulator acts as a control matrix, which can selectively transmit each of the beams independently. The spatial light modulator receives its input from a commercial workstation, which acts as the system host.

The beams that pass through the spatial light modulator are collected by a second train of anamorphic lenses and delivered to a 128-element linear array of avalanche photodiodes. Light from each of the 128 columns in the spatial light modulator is delivered to a specific photodiode.

DOC II successfully demonstrated the feasibility of optical computing concepts, by performing general purpose computations at a rate of 8×10^{11} bit instructions per second, and by performing database text searching. The device operated with a gate switching energy of less than 4 fJ, a value far below that associated with semiconductor devices.

This testbed was not a fully optical computer: the control matrix was an acoustooptic device, operated by electrical input. Substantial further advances will be required to develop all-optical processing.

The status of digital optical computing lags well behind that of its electronic counterpart. It is basically still in a research stage. Significant advances in the development of large arrays of high-density, high-speed optical gates will be required to make optical computing a viable alternative to electronic computing. There are numerous research programs working to accomplish this. It appears likely that optical computing will become practical in the future, especially for applications that can exploit its massively parallel nature, like image processing.

D. Laser Graphics

A wide variety of laser applications have developed in the fields of the graphic arts. These applications include methods to record, reproduce, and transmit graphic and textual information. There have been many applications of laser graphics in the publishing industry, including phototypesetting and platemaking. In what follows, we shall describe a few examples of such systems.

Laser printers, which convert pages of text or graphics stored in a computer to printed material, could also be considered as an example of laser graphics. This application will be described in a later section.

A generalized laser graphic system is illustrated in Figure 24-15. The laser beam is scanned over a recording medium and records a pattern that may represent either printed or pictorial information. The modulator switches the beam on or off as the beam scans to form the desired pattern. The system is controlled by an input signal,

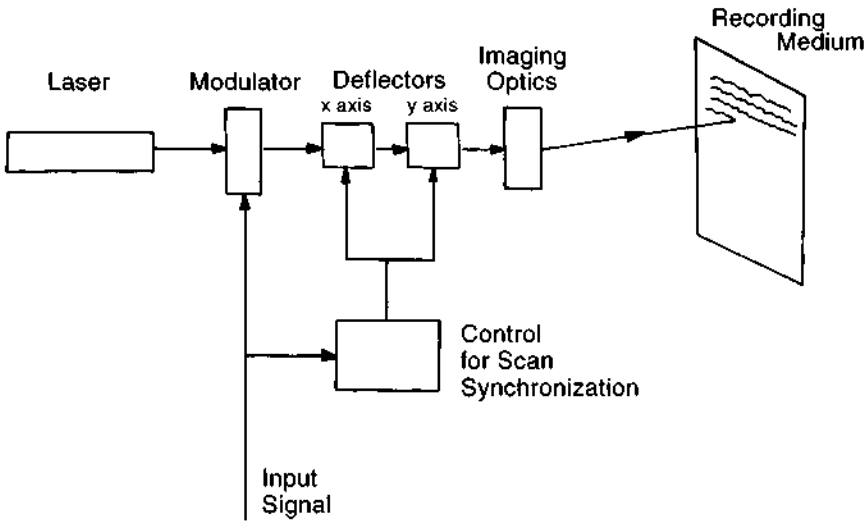


Figure 24-15 A generalized laser graphic system.

which may derive from a variety of sources, according to the application. In a pasteup-to-plate application, it could derive from a detector viewing the reflection from a laser beam scanning a pasteup in synchronism with the scanning of the recording laser. In a direct-to-plate application, it could come directly from a computer in which the information is stored.

A variety of lasers have been employed in laser graphic systems. Historically, helium–neon lasers and air-cooled argon lasers operating at 488 nm have been employed. In more recent times, semiconductor diode lasers, operating in the 700–900 nm range, have become widely used. In applications requiring higher power, diode-pumped Nd:YAG lasers are sometimes used.

A number of different recording materials are used. Many systems employ dry silver halide photographic films, which are developed by heat. Photopolymer materials are also common. Other materials have included polyesters, photoresists, and electrophotographic films. The sensitivity in terms of energy per unit area to expose the different materials varies over orders of magnitude. Their spectral response also varies. Thus, the choice of laser depends strongly on the recording medium chosen. The laser must be able to supply adequate energy per unit area at the right wavelength.

One of the first laser graphics applications in the 1970s allowed scanning of page pasteups in order to produce a printing plate directly. The plate was made right from the pasteup, eliminating intermediate steps and reducing costs. This pasteup-to-plate application has been used in the newspaper industry.

Another early application was laser phototypesetting. Phototypesetting involves photographing characters on film. Phototypesetting machines select, under control of a keyboard, a negative of a desired character. The image of the character is projected onto the recording medium, which later serves as the basis for making the printing surface. The substitution of a laser as the light source in a phototypesetting operation also reduced costs. Sophisticated laser phototypesetting equipment has been used in the newspaper industry. As in nonlaser systems, an operator at a keyboard enters in a character. The electronic logic controls the laser beam to project the image of the character onto the recording medium. Choices of different inputs, like magnetic tape or direct computer control, are also possible.

One significant advantage that laser graphics provides is the ability to deal with photographs, text, and line art simultaneously, on one piece of equipment. All these formats may be projected on the recording medium without having to go through the stage of manual cutting and pasting. Because of this ability, laser graphic systems are sometimes called imagesetters. This capability is being utilized in modern direct-to-plate systems in which printing plates are exposed directly by the imagesetter to form the integrated pages of text and illustrations under the control of the computer that has stored the information. This procedure eliminates several steps in the generation of printing plates and offers savings in both cost and time.

In a recent example [2], a laser imagesetter allowed the publishing, for the first time, of a color magazine by direct writing of images on offset printing plates without use of film. The imagesetter employed a 488 nm argon laser and exposed the

printing plates using a raster scan. Both polymer-based and silver halide plates were used. The direct-to-plate system offered cost reductions through elimination of the film and processing costs.

In another example, a computer-to-plate system uses a series of laser diodes, 24 or 32, depending on the size of the plate. The laser light is brought to the plate by a fiber optic delivery system. The laser beams are scanned horizontally under the control of a computer to form the desired images on the plates.

Dry plates contain three layers, a top layer that repels ink, a middle layer that absorbs light, and a base layer that accepts ink. The laser light is absorbed in the middle layer, causing it to ablate. The ablation process also removes the top layer. This exposes the base layer, in the form of the desired image. This procedure directly produces the plates for printing, with the image formed on an ink accepting layer.

The procedure is done without use of photographic film, film processing, or photographic chemicals. The system offers increased productivity and lowered costs compared with conventional platemaking processes.

Many different laser graphic systems have been developed, using different facets of laser technology and offering advantages for a variety of applications. The ability of laser-based systems to eliminate intermediate steps and to expose a printing plate ready for press, directly from computer input, is an important advance for the printing industry. Applications for sophisticated laser graphic systems capable of computer-to-plate processing are continuing to develop.

E. Consumer Products

For an extended period after the advent of the laser in 1960, most laser applications were in relatively high technology areas, with only a small number of people, mostly scientists, engineers, and technicians, directly involved with their use. The situation has changed dramatically. Laser technology has provided a number of widely used consumer products, which have become very familiar and are used daily by large numbers of nontechnical people. We briefly describe some of the most common of these consumer products.

1. COMPACT DISCS

Optical disk technology has become familiar through the use of prerecorded compact discs (CDs). The largest application has been audio CDs, used largely for the playing of music. Figure 24-16 shows a schematic configuration representative of CD players. The digital information on such discs is retrieved by reading the digital pattern of reflectivity prerecorded in tracks on the spinning disc. The clocking and tracking functions are indicated. Clocking, which refers to the timing of the scanning of the beam along the track of bits, may be controlled by observing the times at which the signals from the bits are observed. Tracking, which refers to keeping the beam accurately aligned to the track of bits, may be performed by deliberately

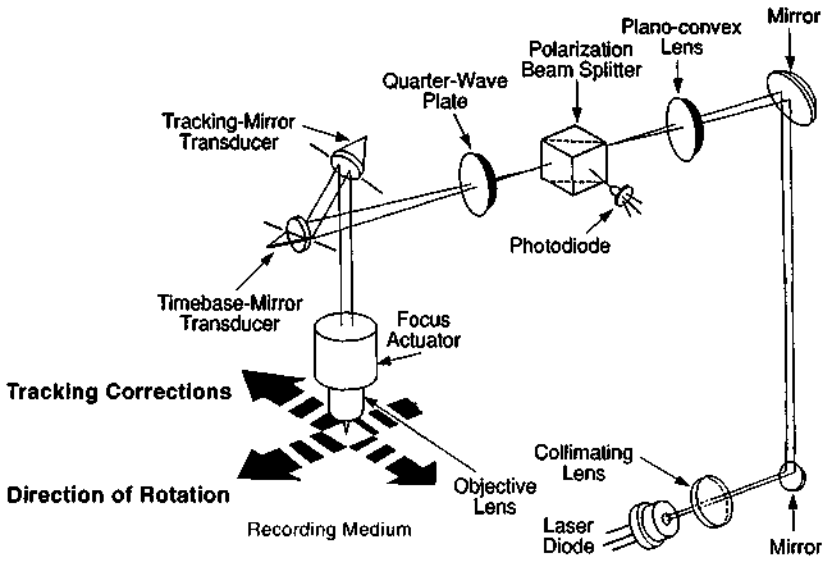


Figure 24-16 Representative configuration for a compact disc player.

offsetting two bits from the track and observing the change in signal as the beam passes them. Focus control may be performed in the same way as was described for Figure 24-6.

The reproduction of the sound recorded in the digital format has very high fidelity. Because there is no contact between the read head and the disc, there is no background noise. Because the read head and the disc surface are some distance apart (around 1 mm), the tolerances on the system are relieved. The presence of small imperfections on the surface of the disk does not affect the reproduction, because they are out of focus for the laser light.

The lasers used for audio CDs have been AlGaAs diodes, operating at a wavelength of 780 nm. This has been the largest single application for lasers. Tens of millions of lasers are manufactured annually for this application.

In addition to their use in audio reproduction, laser disc systems are developing for video applications. Video players capable of playing full-length motion pictures stored on a 5.25 in. disc are becoming available.

Another established use is the compact disc read-only memory (CD-ROM). This has become widely used in conjunction with personal computers.

2. *LASER PRINTERS*

Laser printers have become another familiar laser-based consumer product, often used in conjunction with personal computers. Their principle of operation involves electrophotography, also called xerography, the same process that is used in photocopy machines. The electrophotographic process is illustrated in Figure 24-17. A

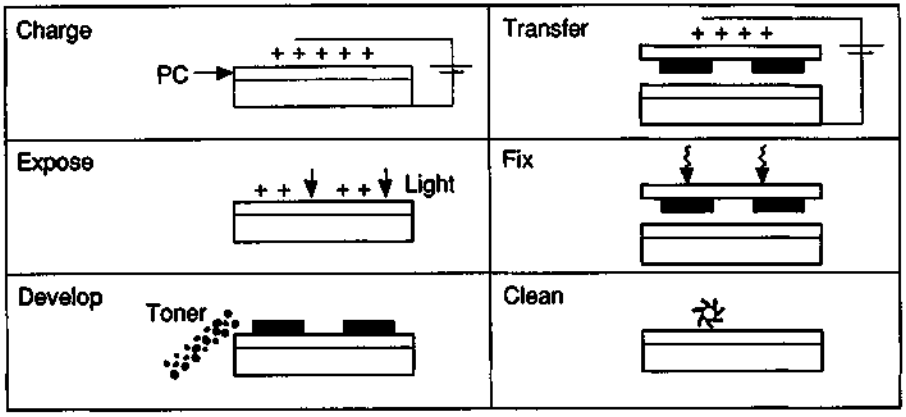


Figure 24-17 Sequence of steps in electrophotography. PC denotes the photoconductive layer.

photoreceptive surface, containing a layer of photoconductive material, is charged uniformly by ions from a corona discharge. The surface is often the surface of a rotating drum. After the charging step, the surface is exposed by scanning the laser beam across it. The intensity of the beam is varied to produce a replica of the image that is desired. The pattern of light intensity reproduces the pattern of the text or graphics that is to be printed. The photoconductive layer becomes conducting in regions illuminated by the laser. This allows the electrostatic charge to move and creates a replica of the pattern in electric charge on the surface.

In the next stage, dry, powdery particles called toner come into contact with the charged surface and adhere to the charged areas. This develops the image as a pattern formed by the toner. Then, the toner particles are transferred to paper by application of the corona discharge again. The toner on the paper is heated to fuse the toner and fix the image on the paper. Finally, the photoreceptor surface is cleaned, to remove residual toner and to prepare the photoreceptive surface for another cycle.

This sequence is essentially the same as that used in photocopy machines, except that in photocopiers, the pattern in the exposure step is obtained from projection of an image of the text or graphics that is to be copied. In the laser printer, the material is generated by a computer and sent to the printer, which stores the data in its memory, composes one page of material at a time, and controls the scanning of the laser beam to produce the desired image.

A possible configuration for a laser printer is shown in Figure 24-18. The figure indicates use of a semiconductor diode laser source. The first laser printers used helium–neon lasers, but in more recent times, the use of visible diode lasers has become common.

Laser printers provide quality printing with sharp images and consistent black tones. They have become widely used for printing of reports and letters and for desktop publishing.

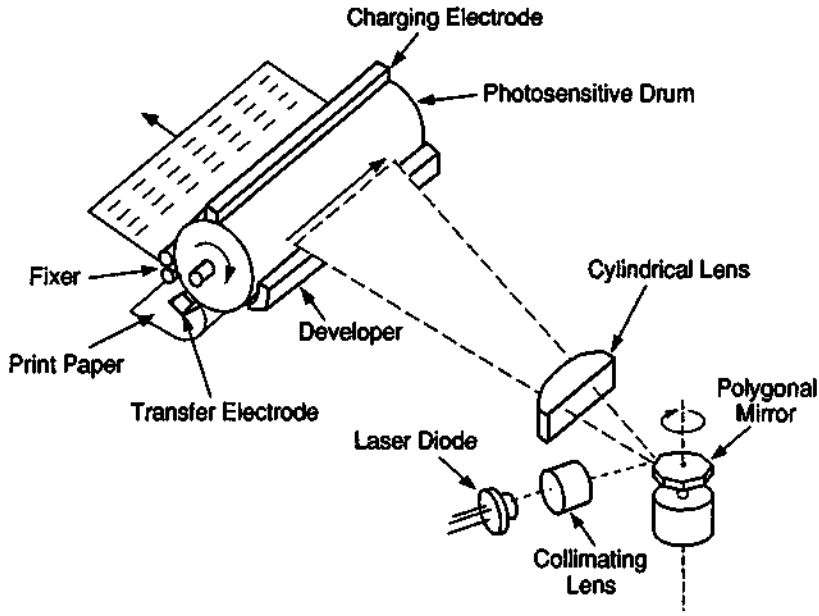


Figure 24-18 Laser printer configuration. (From *Laser Diodes*, Sharp Corp., 1992.)

3. BARCODE SCANNERS

Another arena in which lasers have become commonplace is in barcode scanners, sometimes referred to as supermarkets scanners or as point-of-sale scanners. These devices read the universal product code printed on most items in supermarkets and other stores. The scanner reads the code automatically, looks up the appropriate price in memory, and prints a detailed sales slip. The checker draws the item over a transparent window, without having to punch keys. The scanner contains four basic elements: a low-power laser, a moving mirror to scan the beam in a predetermined pattern, a photodetector to sense variations in the intensity of light reflected from the product code marking, and electronic logic to interpret the signal from the detector. In some devices, the mirror utilizes a holographic optical element, as discussed in Chapter 19.

The code itself is a familiar marking on consumer products. It is a ten-digit code identifying both the supplier and the specific item. It consists of parallel bars of variable width and spacing. When the laser beam scans over this pattern, variations in the intensity of light reflected from light and dark regions allow identification of the item.

The use of a laser source is desirable to provide adequate contrast under conditions of ambient light. The laser beam has low enough power to be eye safe. The lasers used were mainly helium–neon lasers for many years, but visible semiconductor diode lasers are used in newer scanner models.

The projector scans the beam rapidly so as to cover the bar pattern a number of times during the time the item is pulled across the window, so that the item may be identified even if the orientation of the barcode lines is not exact. For the consumer, such scanners offer more rapid checkout and more complete information on sales slips. For the supermarket, scanners offer increased productivity for checkers and the possibility of automatic inventory control.

Barcode scanners have become very common features in stores, libraries, baggage-handling systems, post offices, and so forth, and their use is continuing to increase rapidly. They provide an interface between laser technology and large numbers of persons on a daily basis.

References

[1] R. V. Stone, *Laser Focus World*, p. 77 (August 1994).
[2] S. G. Anderson, *Laser Focus World*, p. 26 (July 1994).

Selected Additional References

R. A. Athale, ed., *Digital Optical Computing*, SPIE Optical Engineering Press, Bellingham, WA, 1990.
D. B. Carlin and D. B. Kay, eds., *Optical Data Storage*, SPIE Proc., Vol. 1663, SPIE, Bellingham, WA, 1992.
M. J. F. Digonnet, ed., *Rare Earth Doped Fiber Lasers and Amplifiers*, Marcel Dekker, New York, 1993.
H. M. Gibbs, *Optical Bistability: Controlling Light with Light*, Academic Press, Orlando, FL, 1985.
Lightwave Technology, Issues of the *AT&T Technical Journal*, January/February 1987 and January/February 1992.
A. B. Marchant, *Optical Recording: A Technical Overview*, Addison-Wesley, Reading, MA, 1990.
A. D. McAuley, *Optical Computer Architecture: The Application of Optical Concepts to Next Generation Computers*, Wiley, New York, 1991.

Epilogue | A Look at the Future

The development of lasers since the first operating devices in 1960 has been phenomenally rapid. They have progressed from fragile laboratory devices to reliable industrial tools. The operating characteristics, including items such as the available range of wavelengths, the peak powers, and the continuous powers, have improved rapidly and continue to do so.

Lasers themselves will continue to develop and proliferate. Rapid and exciting developments in semiconductor diode lasers, diode-pumped solid state lasers, tunable lasers, ultraviolet lasers, and ultrafast lasers are continuing. The ultimate impact of x-ray lasers and of free electron lasers is still speculative, but advances in these technologies are in progress at many laboratories. One need that still remains is a really good high-power visible laser, as good as the helium–neon laser has been, for example. The frequency-doubled Nd:YAG laser partially fulfills that role, but further advances in this area would be desirable.

New applications have been suggested, tested, and brought to practical utilization. A number of laser-based consumer products have become common, including compact disc players, barcode scanners, and laser printers. We may expect consumer applications such as these to increase dramatically.

The impact of lasers on communications has been enormous. The use of lasers and optical fibers in lightwave communications has radically changed the nature of telecommunications and is continuing to expand in other areas of communications. Further advances, including use of erbium-doped fiber amplifiers in all-optical repeater stations and, ultimately, the use of solitons for very wide bandwidth communications, are expected.

Optical data storage applications will continue to grow. The use of optical video discs and personal computer disks should expand rapidly. Other data storage applications will continue to develop.

Optical computing is still mostly in the future, but developmental work is encouraging. Optical computers may begin to replace electronic computers when large arrays of high-density, high-speed optical gates can be fabricated. Digital optical

computing offers the potential advantages of global interconnection, massive parallelism in the processing, extremely low switching energy, and very high speed.

Continued growth and expansion of established industrial applications appears assured. This includes laser material processing, which will continue to replace conventional material processing technologies. This growth will probably not be revolutionary, but it should be steady.

One potential application of lasers is laser-assisted thermonuclear fusion. This represents one of the most exciting possibilities for future laser application. Laser fusion is a form of inertial confinement fusion (ICF). For the fusion reaction to occur effectively, a high-temperature plasma must be confined at a high density for some minimum time. In ICF, the plasma is generated by a very short pulse of energy deposition, in a time so short that the inertia of the material does not allow it to dissipate. ICF competes with magnetic confinement fusion, in which the material is confined by magnetic forces. Devices such as tokamaks are being developed for magnetic confinement fusion. It is too early to be certain whether magnetic confinement or ICF will prove to be the eventual superior technology for power generation.

The prospects for laser-assisted thermonuclear fusion and its eventual importance as a commercial source of power are still uncertain. Its ultimate significance has not yet been completely evaluated. At its outer limit, it could offer the possibility of abundant energy for mankind for all the foreseeable future.

We will conclude with a brief discussion of some of the developments aimed toward laser-assisted thermonuclear fusion. The basic concept involves focusing a high-power laser beam onto a target containing a mixture of deuterium and tritium gases. The mixture is heated to a temperature, around 10^8 K (10 keV). At this temperature, the thermonuclear fusion reaction

$$^2\mathrm{H} + {}^3\mathrm{H} \rightarrow {}^4\mathrm{He} + n \tag{E.1}$$

can occur. There is a net release of energy in this reaction. In experiments to date, the gas mixture is contained in a glass microsphere with diameter around 100 μm. A number of laser beams are incident on the sphere from different directions. The pressure produced by absorption of the light drives an implosion in which a high-pressure shock front travels into the spherical volume. When the implosion reaches the center, the gas is compressed to high density, temperature, and pressure, conditions suitable for the thermonuclear reaction to occur.

Experiments leading toward laser-assisted thermonuclear fusion are being conducted at a number of laboratories worldwide. These experiments have demonstrated the generation of neutrons from reaction (E.1). But they have so far required more energy as input than has been generated by the reaction. One significant milestone is referred to as scientific breakeven. This is the point at which the release of thermonuclear energy equals the laser energy input. This condition has not yet been achieved. The production of a net total gain of energy through generation of enough thermonuclear energy to overcome all the losses and inefficiencies of the process is much further away.

Laser-assisted thermonuclear fusion experiments have been dominated by Nd:glass lasers. The largest such laser has been the so-called Nova laser located at Lawrence Livermore National Laboratory [1]. This system used Nd:phosphate glass, capable of operation at 1.05 μm and at the second and third harmonics (0.53 and 0.35 μm). Coupling of the laser energy to the thermonuclear plasma improves at the shorter wavelengths, so that they are desirable. The Nova laser generated ten beams, which were delivered to the target from ten different directions, so as to drive the implosion fairly uniformly.

Each beam began with a master oscillator, operating at 1.05 μm. The pulse from the master oscillator was amplified by a chain of Nd:glass rods and disks, with increasing diameter, in order to remain below the threshold for laser damage. The final disks had a diameter of 46 cm. The beam line also contained optical isolators to keep energy from being reflected backward, spatial filters to keep the beam profile smooth, and nonlinear optical elements for frequency multiplication.

The Nova laser produced output at 1.05 μm of 125 kJ of laser energy in a 2.5 nsec pulse, and of 80 kJ in a 1 nsec pulse. At 0.35 μm, 45 kJ in a 2 nsec pulse was obtained. Experiments with this laser led to much greater understanding of the physics of the laser interaction with the deuterium–tritium target material.

A proposed next step is the development of a National Ignition Facility, also based on Nd:glass, and leading toward the demonstration of scientific breakeven. As a demonstration of the technology required for the National Ignition Facility, in the mid 1990s, a single-beam experimental laser system is being operated to test the design and performance of the proposed next generation fusion laser. This project, called Beamlet [2], has the goal of demonstrating features of laser science and technology as the basis for the National Ignition Facility.

In the mid 1990s, a single beam produced 125 terawatts (TW) of power in a pulse less than 500 nsec long. At the time, this represented the world's most powerful laser. It was anticipated that the output power could be increased to more than 1 petawatt (10^{15} W, PW).

The National Ignition Facility is planned to have 192 beams, all identical to the Beamlet. These beams should be able to deliver 1.8 MJ of ultraviolet energy to the target in a pulse with duration around 3 nsec.

An ultimate goal is the operation of a demonstration power plant capable of net electrical power generation by the year 2025.

Although much of the research in laser-assisted thermonuclear fusion has used Nd:glass lasers, because of their capability for storage of large amounts of energy and their availability in large size with good optical quality, they will not be the lasers ultimately used for a laser power plant. This is because of the relatively low pulse repetition rate that is possible. This in turn limits the average power available. Development of other lasers, scalable to high values of average power, is proceeding at many laboratories worldwide. Because the coupling of laser energy to the target material improves as the wavelength shortens from infrared to visible to ultraviolet, the laser of choice will probably be ultraviolet. Much of the development of advanced fusion laser candidates centers on excimer lasers, particularly KrF.

Although practical generation of energy by laser-assisted thermonuclear fusion remains well in the future, the program has provided some of the most stressing requirements for laser technology. Many significant advances in lasers have been developed for use in research in this area. These advances have then been utilized in other areas within the laser community. Thus, thermonuclear fusion applications have been an important driving force for all laser technology.

References

[1] C. Bibeau *et al.*, *Appl. Opt.* **31**, 5799 (1992).

[2] J. R. Murray, J. H. Campbell, and D. N. Frank, *Beamlet Project: Technology Demonstration for a National Inertial Confinement Fusion Ignition Facility*, paper CTuC1 at the Conference on Lasers and Electro-Optics, Baltimore, MD, May 2–7, 1993.

Selected Additional References

D. C. Cartwright *et al.*, Excimer Could Be Laser of Choice for Next-Generation ICF Driver, *Laser Focus World*, p. 103 (March 1990).

C. Yamanaka, *Introduction to Laser Fusion*, Gordon and Breach, Reading, UK, 1991.

This Page Intentionally Left Blank

INDEX

A

Absorption, 12
Absorption spectroscopy, 493–495
Acoustooptic effect, 168–171, 175–176, 177, 178, 179, 553, 554
Alexandrite laser, 91, 101, 338
AlGaAs laser, 39, 110–118
AlGaInP laser, 111–112, 133, 136
Alloying, 380–381
American National Standards Institute, 215, 227, 228, 230
Ammonia laser, 527
Amplitude holograms, 452–453, 454, 455
Analog optical computing, 573–579
Angle tracking, 242–243
Angular rotation rate measurement, 287–294
Annealing, 428–430
Argon fluoride laser, 87, 427, 512
Argon laser, 33, 39, 53, 57, 70–74, 230
Atomic vapor laser isotope separation (AVLIS), 523–524, 528

B

Balancing, 413–415
Barcode scanners, 588–589
Beam divergence, 36–40
Beam expanders, 39–40
Beam homogenizer, 190–191
Beam modulation telemetry, 257, 270–274
Beam profiling, 185–187
Beamlet, 592
Beat production, 251–252
Birefringence, 6–7
Brewster's angle, 6

C

Carbon dioxide laser, 33, 39, 53, 57, 75–85, 230
Carbon monoxide laser, 84–85, 373, 541
Cavity dumping, 52–53
Chemical applications, 500–501, 510–528
Chemical laser, 136–137
Chemical vapor deposition, 430–432
Cladding, 381–382
Coherence, 53–58, 253–255
Collimation, 36–40
Co:Magnesium fluoride laser, 91
Compact discs, 585–586
Construction applications, 244–246
Consumer products, 585–589
Copper vapor laser, 88, 338, 371, 523
Coumarin 6 laser, 126, 127
Coumarin 110 laser, 126, 127
Cr:LiCAF laser, 102, 136
Cr:LiSAF laser, 102, 136
Cr:YAG laser, 91
Cutting, 339, 395–409, 416–417, 422, 423
Cylindrical form measurement, 310–311

D

DCM special laser, 126–127
Deep penetration welding, 353–364
Defect detection, 311–313
Deflectors, 171–176, 553–554
Detectivity, 154, 158, 163
Detectors, 152–165
Deuterium fluoride laser, 137
DIAL spectroscopy, 507–508, 509
Diffraction, 7, 8
Diffraction limit, 59–62
Diffraction measurements, 295–297
Diffractive optics, 311
Digital optical computing, 580–583
Diode lasers, 102–120. *See also* Semiconductor laser; AlGaAs laser
Diode pumped solid state laser, 91, 95–96, 133–136
Distance measurement, 257–276
Distributed Bragg reflector laser, 119, 552–553, 556, 557, 558
Doppler effect, 252–253, 497–498
Doppler-free spectroscopy, 494, 497–499
Double exposure holographic interferometry, 468, 470–471, 479
Drilling, 339, 387–395, 421–422, 424–425
Dye laser, 53, 120–129

E

Electromagnetic radiation, 1–4
Electron-beam-controlled laser, 84
Electrooptic effect, 165–168, 171, 174–175, 177
Electrophotography, 586–587
Energy levels, 9–11, 14, 15
Environmental monitoring, 501–509, 543–544
Erbium doped fiber amplifiers (EDFA), 540, 565, 566
Erbium laser, 91
Er:YAG laser, 541
Etalon, 34–35, 211, 212
Excimer laser, 86–87. *See also* KrF laser
Eye damage, 216–222

F

Fiber amplifiers, 540
Fiber loss, 560–563
Fiber optic gyroscope (FOG), 294, 543
Fiber optic sensors, 542–544
Fiber optics, 530–544, 560–566
Flashlamp, 18, 19
Fluid flow measurement, 278–284
Fluorescence, 11, 12
Focusing of laser light, 59–63, 335
Four level systems, 17, 18
Fourier transform hologram, 441–442
Free electron laser, 137–139, 142, 143
Fresnel hologram, 440

G

GaAs laser, 114–117, 535
Gabor, D., 440, 480, 481
Galilean beam expander, 39
Gallium nitride laser, 133
Gas dynamic laser, 83–84
Gas lasers, 66–88. *See also* Helium–neon laser
Gaussian beam, 36–38, 41–43, 59, 232
Gaussian line, 14, 15
Glazing, 380
Gold vapor laser, 88
Graded index fiber, 533

H

Hardening, 373–380, 382
Heat treating, 339, 373–380, 382
Helium–cadmium laser, 73–75, 341
Helium–neon laser, 33, 39, 53, 57, 67–70, 230
Heterostructure (heterojunction) laser, 114–117
High temperature superconductor, 416
Holmium laser, 91, 100
Holographic applications, 464–489
Holographic data storage, 482, 571–573
Holographic interferometry, 464–481, 487–489
Holographic optical elements, 486–487, 588

Holographic optical memory, 571–573
Holographic recording materials, 458–462
Holographic stereogram, 484–485
Holography, 437–462
Homogeneous line, 14, 15
Homostructure (homojunction) laser, 114
Hydrogen chloride laser, 137
Hydrogen cyanide laser, 85, 86
Hydrogen fluoride laser, 137

I

Infrared fibers, 540–542
Infrared materials, 151–152
InGaAs laser, 112
InGaAsP laser, 111–112, 142, 535, 563
Inhomogeneous line, 14–15
Injection seeding, 35, 184–185
Integrated optics, 546–558
Interferometric distance measurement, 257–269, 276
Ionized gas lasers, 70–75. *See also* Argon laser
Isotope separation, 518–528

K

Keplerian beam expander, 39
Keyholing, 319, 353, 354, 355, 360
Krypton laser, 73–74
KrF laser, 86–87

L

Lamb dip, 36
Laser. *See* specific laser types, e.g., Nd:YAG laser
Laser-assisted thermonuclear fusion, 591–593
Laser Doppler displacement, 257, 270
Laser graphics, 583–585
Laser printers, 586–588
Laser-supported absorption wave, 327, 328, 329, 332, 333, 334
Lead selenide sulfide laser, 502
Leith, E. N., 440
LIDAR (optical radar), 274, 502, 503, 509
Lightwave communications, 559–566
Linewidth, 31–36
Longitudinal modes, 24, 31–36, 55–56, 456–457
Lorentzian line, 14, 15
Losses in optical fibers, 533–537

M

Marking, 339, 411–414, 423–424
Maximum permissible exposure, 227, 228, 229
Mechanical deflectors, 171, 172–174
Metal vapor laser, 87–88
Michelson interferometer, 249–251
Mirrors, 144, 147
Mode hopping, 108–109
Mode locking, 51–52
Modulators, 165–171, 551–552
Molecular lasers, 75–86. *See also* Carbon dioxide laser
Momentum transfer, 330
M squared, 43–44
Multimode fibers, 532, 533, 564
Mutual coherence function, 54–57

N

National Ignition Facility, 592
Nd:glass laser, 33, 39, 53, 57, 96–98
Nd:GSGG laser, 91
Nd:YAG laser, 39, 53, 57, 89–96, 133–136, 230
Nd:YLF laser, 91, 98–99
Neutral gas laser, 67–70. *See also* Helium–neon laser
Nitrogen laser, 84–85, 123, 125, 512, 518
Nitrous oxide laser, 85
Noise equivalent power, 154
Nonlinear optics, 179–182
Normal pulse operation, 47–48
NOVA laser, 592
Numerical aperture, 536, 537, 539

O

Off-axia holography, 438, 440
Optical bistability, 581–582
Optical data processing, 573–583
Optical data storage, 566–573
Optical isolators, 182–183
Optical parametric oscillator, 140–141, 142, 143, 511, 514
Optical pumping, 18–20
Optical radar. *See* LIDAR
Optical tables, 187–188
Optical waveguides, 547–551
Oxygen–iodine laser, 137

P

Paint stripping, 415
Particle diameter measurement, 307–309
Partical image velocimetry, 283–284
PbEuSe laser, 112
PbSe laser, 112
PbSnSe laser, 112
Phase hologram, 452–453, 454, 455
Photoconductive effect, 154, 158, 162
Photodiodes, 155–159
Photoemissive effect, 154, 159–161
Photolithography, 426–427
Photomask repair, 434–435
Photomultiplier, 161
Photovoltaic effect, 154, 155–159
Photon detectors, 153, 154, 158, 159
Picosecond spectroscopy, 500
Plasma shielding, 327, 328, 329
Polarizers, 149–151
Polyphenyl 2 laser, 126, 127
Population inversion, 15–18
Position-sensitive detectors, 233–235
Product dimension measurement, 303–305
Profile measurement, 297–303
Pulsed range finding, 257, 274–275
Pyridine 2 laser, 126, 127

Q

Q-switching, 48–51, 176–179

R

Radiance, 58–59
Radiometric units, 58–59
Rainbow holography, 484–485
Raman shifters, 183–184
Raman spectroscopy, 494–497, 502–504, 509
Rapid prototyping, 341
Read-only memories, 567, 585–586
Real-time holographic interferometry, 465–469, 479
Reflection holograms, 453, 454, 455
Reflectivity, 320, 321, 322
Remote sensing, 501–509
Resonance fluorescence, 503–506, 509
Resonant cavity, 22–29
Responsivity, 153–154, 157
Rewritable memories, 568–569
Rhodamine 6G laser, 121, 125, 126, 127
Rhodamine 110 laser, 126, 127
Rhodamine 700 laser, 126, 127
Ring laser gyroscope, 287–294
Ruby laser, 33, 39, 53, 57, 99–100

S

Sagnac effect, 288
Scribing, 339, 409–411, 422–423
Seam welding, 345–364, 367–369
Self-electrooptic effect device (SEED), 582
Semiconductor laser, 102–120, 132–133, 203–204, 212–213. *See also* AlGaAs laser
Single-mode fibers, 532, 533, 564
Skin hazards, 222–223
Solid state laser, 89–102. *See also* Nd:YAG laser
Solitons, 566
Spatial light modulators, 189–190, 571
Speckle pattern, 57–58
Spectroscopic applications, 491–509
Spot welding, 363–367
State-to-state chemistry, 518
Stereolithography, 341
Stilbene 1 laser, 126, 127
Stilbene 3 laser, 126, 127
Stimulated emission, 12, 16
Strain measurement, 309–310

Styrl 9 laser, 126, 127
Sulforhodamine B laser, 125, 126, 127
Surface finish measurement, 305–30
Surface position measurement, 297–303
Surface velocity measurement, 284–287

T

TEA carbon dioxide laser, 82–83
TEM modes, 40–46
Thermal detectors, 153, 161–163
Thermal diffusivity, 316, 317, 318
Thermal time constant, 316, 317
Thermonuclear fusion, 591–593
Thick holograms, 451–452, 454, 455
Thin film deposition, 415–416
Thin holograms, 451–452, 454, 455
Three level systems, 17, 18
Threshold for laser operation, 16, 17
Time-average holographic interferometry, 470, 472–474, 479
Ti:sapphire laser, 101, 102
Tm:YAG laser, 91
Tooling, 235–242
Transmission holograms, 453, 454, 455
Transverse modes, 40–46
Trimming, 419–421
Tunable lasers, 141–143, 492–493

U

Upatnieks, J., 440

V

Velocity measurement, 278–287
Vibration measurement, 310
Vibronic solid state laser, 100–102, 143

W

Water vapor laser, 85
Waveguide carbon dioxide laser, 83
Welding, 339, 343–371
Write-once read-mostly (WORM) memories, 567–568

X

Xenon chloride laser, 87
Xenon fluoride laser, 87
Xenon laser, 75, 420
X-ray laser, 139–140, 440, 482

ISBN 0-12-583961-8