AGARDograph No. 296

Laser Applications in Flow Diagnostics

NORTH ATLANTIC TREATY ORGANIZATION
AGARDograph No.296

LASER APPLICATIONS IN FLOW DIAGNOSTICS

by

J.D. Trolinger
Chief Scientist
Spectron Development Laboratories, Inc.
Costa Mesa, California 92626-1439
USA

Edited by

J.J. Ginoux
Director
Von Kármán Institute for Fluid Dynamics
Chaussée de Waterloo 72
B-1640 Rhode-Saint-Genèse
Belgium

This AGARDograph has been produced at the request of the Fluid Dynamics Panel of AGARD.
THE MISSION OF AGARD

According to its Charter, the mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;
- Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Exchange of scientific and technical information;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced directly from material supplied by AGARD or the author.

Published October 1988
Copyright © AGARD 1988
All Rights Reserved
ISBN 92-835-0480-1

Printed by Specialised Printing Services Limited
40 Chipwell Lane, Loughton, Essex IG10 3TZ

ii
PREFACE

Over 15 years ago I began the preparation of AGARDograph Number 186, published in 1974, the precursor with the same title as this publication. The first publication provided a summary of the state of the art in the application of lasers in the field of flow diagnostics. Example applications were provided from a number of NATO laboratories, including many from the author's laboratory at the US Air Force Arnold Engineering Development Center.

The present project began as simply an update and expansion of the earlier publication. I quickly came to the realization that the field has expanded and matured so much that a virtually complete rewrite was necessary. The resulting volume contains more new material than repeated and updated material. New hardware and technology have evolved, including detector arrays, small inexpensive personal computers, powerful software, a wide array of spectroscopic techniques, and many other new techniques.

Literature surveys and visits to various laboratories produced several thousand excellent publications of techniques and applications. One of the difficulties in compiling this work was the need to limit its length and to choose only representative applications from the many fine examples I saw during the laboratory visits. I have added one entirely new chapter to the volume—laser spectroscopy—and retained the remaining four which cover (1) a brief introduction to coherent optics and lasers, (2) conventional flow visualization techniques that can benefit from the use of lasers as a source of light, (3) flow visualization and particle field holography, and (4) laser velocimetry.

Major differences in the subject of laser flow diagnostics today are that it is a mature technology, being applied routinely in hundreds of laboratories, and many varieties of laser diagnostics systems are commercially available. This does not mean that the field has stagnated or even reached a plateau in producing new techniques. For example, laser fluorescence and molecular velocimetry made possible by excimer lasers is an exploding technology that will have critical application in high speed flows owing to revived interest in hypersonic aircraft.

I fully anticipate the necessity to rewrite this again before the year 2000.

***


Le présent document fut à l'origine une simple mise à jour destinée à apporter un complément d'informations à l'AGARDographie 186. Or, j'ai conclu très rapidement que les connaissances dans ce domaine avaient considérablement évolué, atteignant un degré de sophistication tel que je serais amené à entreprendre une refonte quasi-totalement du document original. L'AGARDographie qui en résulte comprend une part plus importante d'éléments nouveaux que d'éléments repris à jour. Du nouveau matériel a fait son apparition, grâce aux technologies nouvelles; notamment: des dispositifs de détection, des petits ordinateurs personnels de coût modique, des logiciels performants, ainsi que toute une gamme de techniques spectroscopiques et d'autres techniques nouvelles. Les visites que nous avons pu effectuer dans différents laboratoires, ainsi que la synthèse de la littérature que nous avons faite nous ont permis de disposer de plusieurs milliers de publications d'un très haut niveau concernant les technologies en question.

Cependant, l'une des difficultés que j'ai rencontrée lors de l'élaboration du présent document a été de limiter son volume en ne choisissant que des applications représentatives parmi les nombreux exemples présentés, au demeurant tous d'une qualité remarquable. J'ai ajouté un nouveau chapitre au présent volume, concernant la spectroscopie laser, qui vient compléter les quatre chapitres du document original. (1) Introduction à l'optique cohérente et aux lasers; (2) techniques classiques de visualisation de l'écoulement susceptibles de bénéficier de l'emploi de lasers comme source lumineuse; (3) visualisation de l'écoulement et holographie du champ des particules; (4) vélocimétrie laser.

Aujourd'hui, le diagnostic de l'écoulement au laser est une technologie dont l'efficacité n'est plus à démontrer à laquelle des centaines de laboratoires font appel tous les jours. D'autre part, est apparu sur le marché un grand choix de systèmes de diagnostic au laser.

Cela ne veut pas dire que les travaux de recherche dans ce domaine stagnent pour autant, le seul d'innovation étant loin d'être atteint. A titre d'exemple, nous citerons la fluorescence laser et la vélocimétrie moléculaire, tributaires toutes les deux de la technique des lasers à excimaire; ils s'agît là d'une technologie en plein essor, qui est appelée à jouer un rôle déterminant dans le domaine de l'écoulement à grande vitesse, suite au regain d'intérêt pour les avions hypersoniques.

En conclusion, il me semble pertinent de signaler qu'une nouvelle refonte document s'imposera avant l'an 2000.
# CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER 1 - INTRODUCTION TO LASERS AND LIGHT</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMMARY</td>
<td>1</td>
</tr>
<tr>
<td>1.0 BACKGROUND</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Description of Light</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 Waves and Wavefronts</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2 Properties</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2.1 Polarization</td>
<td>2</td>
</tr>
<tr>
<td>1.1.2.2 Amplitude and Intensity</td>
<td>2</td>
</tr>
<tr>
<td>1.1.2.3 Coherence</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Interference</td>
<td>3</td>
</tr>
<tr>
<td>1.2.1 Plane Waves</td>
<td>3</td>
</tr>
<tr>
<td>1.2.2 Spherical Waves</td>
<td>5</td>
</tr>
<tr>
<td>1.2.3 The Moire Analogy</td>
<td>6</td>
</tr>
<tr>
<td>1.3 Interaction of Light with Matter and Boundaries</td>
<td>7</td>
</tr>
<tr>
<td>1.3.1 Diffraction</td>
<td>7</td>
</tr>
<tr>
<td>1.3.2 Reflection and Refraction</td>
<td>9</td>
</tr>
<tr>
<td>1.3.3 Propagation</td>
<td>10</td>
</tr>
<tr>
<td>1.4 Control of Light</td>
<td>13</td>
</tr>
<tr>
<td>1.4.1 Lenses and Mirrors</td>
<td>13</td>
</tr>
<tr>
<td>1.4.2 Scanners, Deflectors, Modulators, and Shutters</td>
<td>15</td>
</tr>
<tr>
<td>1.4.3 Holographic Optical Elements</td>
<td>16</td>
</tr>
<tr>
<td>1.4.4 Spatial Filtering</td>
<td>16</td>
</tr>
<tr>
<td>1.4.5 Illuminating an Object Field</td>
<td>17</td>
</tr>
<tr>
<td>1.5 Lasers</td>
<td>18</td>
</tr>
<tr>
<td>1.5.1 How They Work</td>
<td>18</td>
</tr>
<tr>
<td>1.5.1.1 Stimulated Emission</td>
<td>18</td>
</tr>
<tr>
<td>1.5.1.2 Population Inversion</td>
<td>18</td>
</tr>
<tr>
<td>1.5.1.3 Amplification</td>
<td>18</td>
</tr>
<tr>
<td>1.5.1.4 Feedback-The Oscillator</td>
<td>18</td>
</tr>
<tr>
<td>1.5.2 Coherence Improvement and Control</td>
<td>20</td>
</tr>
<tr>
<td>1.5.2.1 Frequency Broadening</td>
<td>20</td>
</tr>
<tr>
<td>1.5.2.2 Transverse Electromagnetic Modes (T.E.M.)</td>
<td>20</td>
</tr>
<tr>
<td>1.5.2.3 Longitudinal Modes</td>
<td>20</td>
</tr>
<tr>
<td>1.5.2.4 Single Frequency Lasers</td>
<td>22</td>
</tr>
<tr>
<td>1.5.2.5 Stability and Tuning</td>
<td>22</td>
</tr>
<tr>
<td>1.5.2.6 Q-Switching</td>
<td>23</td>
</tr>
<tr>
<td>1.5.3 Commercially Available Lasers</td>
<td>24</td>
</tr>
<tr>
<td>1.5.3.1 Summary</td>
<td>24</td>
</tr>
<tr>
<td>1.5.3.2 Solid State Lasers</td>
<td>24</td>
</tr>
<tr>
<td>1.5.3.3 Gas Lasers</td>
<td>26</td>
</tr>
<tr>
<td>1.5.3.4 Diode Lasers</td>
<td>27</td>
</tr>
<tr>
<td>1.5.3.5 Metal Vapor Lasers</td>
<td>28</td>
</tr>
<tr>
<td>1.5.3.6 Excimer Lasers</td>
<td>28</td>
</tr>
<tr>
<td>1.5.3.7 Dye Lasers</td>
<td>28</td>
</tr>
<tr>
<td>1.5.3.8 T.E.A. Lasers</td>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 2 - USING LASERS WITH CONVENTIONAL OPTICAL TECHNIQUES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0 BACKGROUND</td>
<td>29</td>
</tr>
<tr>
<td>2.1 Laser Photography and Videography</td>
<td>29</td>
</tr>
<tr>
<td>2.1.1 Photography of Hypervelocity Objects</td>
<td>30</td>
</tr>
<tr>
<td>2.1.2 Videography of High-Speed Events</td>
<td>32</td>
</tr>
<tr>
<td>2.2 Structured Light Methods</td>
<td>36</td>
</tr>
<tr>
<td>2.2.1 Sheet-of-Light Flow Visualization</td>
<td>36</td>
</tr>
<tr>
<td>2.2.2 Particle Image Displacement Velocimetry (P.I.D.)</td>
<td>37</td>
</tr>
<tr>
<td>2.2.3 Deflectometry and Hartmann Methods</td>
<td>40</td>
</tr>
<tr>
<td>2.3 Shadowgraphy and Schlieren Methods</td>
<td>42</td>
</tr>
<tr>
<td>2.3.1 Space Shuttle Schlieren System</td>
<td>43</td>
</tr>
</tbody>
</table>
## CHAPTER 4 - LASER VELOCIMETRY

### 4.0 BACKGROUND

<table>
<thead>
<tr>
<th>4.1 Laser Doppler Methods</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.1 Theoretical Basis</td>
<td>100</td>
</tr>
<tr>
<td>4.1.1.1 Doppler Theory</td>
<td>100</td>
</tr>
<tr>
<td>4.1.1.2 Fringe Models</td>
<td>102</td>
</tr>
<tr>
<td>4.1.1.3 The Moire Analogy</td>
<td>104</td>
</tr>
<tr>
<td>4.1.2 Measuring the Doppler Frequency</td>
<td>104</td>
</tr>
<tr>
<td>4.1.2.1 Direct Optical Frequency/Wavelength Measurement</td>
<td>104</td>
</tr>
<tr>
<td>4.1.2.2 Heterodyning Methods</td>
<td>105</td>
</tr>
<tr>
<td>4.1.2.3 Signal Analysis and Interpretation</td>
<td>107</td>
</tr>
<tr>
<td>4.1.3 Optical Geometries</td>
<td>110</td>
</tr>
<tr>
<td>4.1.3.1 Systems</td>
<td>110</td>
</tr>
<tr>
<td>4.1.3.2 Two Component Dual Scatter Systems</td>
<td>112</td>
</tr>
<tr>
<td>4.1.3.3 Three Component Dual Scatter Systems</td>
<td>112</td>
</tr>
<tr>
<td>4.1.3.4 Moving Fringes</td>
<td>113</td>
</tr>
<tr>
<td>4.1.3.5 Fiber Optics</td>
<td>113</td>
</tr>
<tr>
<td>4.1.3.6 Reference Scatter and Other Direct Doppler Systems</td>
<td>114</td>
</tr>
<tr>
<td>4.1.4 Electronic Signal Processors</td>
<td>115</td>
</tr>
<tr>
<td>4.1.4.1 Oscilloscopes</td>
<td>115</td>
</tr>
<tr>
<td>4.1.4.2 Spectrum Analyzers</td>
<td>115</td>
</tr>
<tr>
<td>4.1.4.3 Frequency Trackers</td>
<td>115</td>
</tr>
<tr>
<td>4.1.4.4 Electronic Counters</td>
<td>115</td>
</tr>
<tr>
<td>4.1.4.5 Simultaneity Test</td>
<td>115</td>
</tr>
<tr>
<td>4.1.4.6 Transient Recorders</td>
<td>116</td>
</tr>
<tr>
<td>4.1.4.7 Correlators</td>
<td>116</td>
</tr>
<tr>
<td>4.1.5 Applications</td>
<td>116</td>
</tr>
<tr>
<td>4.1.5.1 Wind Tunnels - ONERA (France)</td>
<td>117</td>
</tr>
<tr>
<td>4.1.5.2 Shock Tubes</td>
<td>119</td>
</tr>
</tbody>
</table>
### 4.1.5.3 Long Range Systems

- Page: 119

### 4.1.5.4 Fiber-Optical Systems

- Page: 120

### 4.5.5.5 Turbomachinery and Internal Combustion Engines

- Page: 121

### 4.1.5.6 Spatial Correlation Measurements

- Page: 122

### 4.1.5.7 Flow Seeding

- Page: 122

#### 4.2 Transit (Two-Spot) Anemometry (Velocimetry)

1. **Theoretical Basis**
   - 4.2.1.1 Concept
   - 4.2.1.2 Sample Volume
   - 4.2.1.3 Nature of the Signal
   - Enhancing the Signal with Filters and Correlators
   - Directional Sensing and Turbulence Measurement

2. **Optical System**
   - 4.2.2.1 Beam Splitting
   - 4.2.2.2 Stoppers
   - 4.2.2.3 Spot Intensity and Aberrations
   - 4.2.2.4 Spot Rotation
   - 4.2.2.5 Detection

3. **Signal Processing and Interpretation**
   - Broadening
   - Data Rate
   - Time Measurement
   - Data Analysis

4. **Comparison of LTA to LDV**

### Applications

- Turbomachinery
- Wind Tunnels

### 4.2.6 Two-Line Anemometers

---

### CHAPTER 5 - LASER SPECTROSCOPY AND SCATTERING METHODS

#### 5.0 INTRODUCTION AND BACKGROUND

- Page: 139

#### 5.1 Rayleigh Scattering

1. **Principle of Operation**
2. **Analytical Description**

#### 5.2 Absorption/Emission Spectroscopy

1. **Principle of Operation**
2. **Analytical Description**

#### 5.3 Laser Induced Fluorescence

1. **Principle of Operation**
2. **Analytical Description**
3. **Special Considerations**

#### 5.4 Raman Spectroscopy

1. **Principle of Operation**
2. **Technique Description**

#### 5.5 Coherent Raman Spectroscopy

1. **Principle of Operation**

#### 5.6 Application Examples

1. **Rayleigh Scattering for Measurement of Density in the Shuttle Reentry Path**
2. **Laser Induced Fluorescence Measurements of Oxygen in Supersonic Flows**
3. **Wind Tunnel Applications of Coherent Raman Scattering**

---

### APPENDIX A - MOIRÉ PATTERN MASTERS

- Page: 153

### APPENDIX B - REFERENCES

- Page: 156

### APPENDIX C - ADDITIONAL REFERENCES

- Page: 167

### ACKNOWLEDGEMENTS

- Page: 178
LASER APPLICATIONS IN FLOW DIAGNOSTICS

by
James D. Trolinger
Chief Scientist
SPECTRUM DEVELOPMENT LABORATORIES, INC.
Costa Mesa, California, U.S.A.

SUMMARY
This is an expansion and update of AGARDograph No. 186, Laser Instrumentation for Flow Field Diagnostics, which was published in 1974. A brief introduction to coherent optics and lasers is given to provide a review and common language to the reader. The material is presented at a level to be understood by engineers who are working in the aerospace field and who are interested in the experimental application of lasers in research and development. An emphasis is placed on applications in aerodynamics, and examples have been selected from a number of laboratories in the NATO countries visited by the author while preparing the material. The applications include holography, laser velocimetry, spectroscopy, and the use of lasers simply as efficient light sources in conventional measurement methods.

CHAPTER 1
INTRODUCTION TO LASERS AND LIGHT

1.0 BACKGROUND
An elegant and accurate theoretical description of coherent light existed many years before such light was available to experimentalists, especially those who were involved in practical applications. The advent of the laser in the early sixties transformed an essentially academic theory into a powerful and practical tool, which could be easily applied to direct the path to a vast number of applications. It was, perhaps, this combination of the availability of a powerful new experimental tool, and a highly refined theory with which to describe it, that led to a virtual explosion of the field of applied optics. The growth was accelerated not only in experimental fields but also in theoretical optics which began to feed on the ever increasing number of experimental anomalies. The large number of practical applications, moreover, provided both interest and support for a significant increase in the number of people working and contributing to the field. In this chapter we will provide a brief summary of coherent optics and lasers for the purpose of providing the reader with the necessary language and basics to understand later chapters.

1.1 Description of Light

1.1.1 Waves and Wavefronts
Light is a transport phenomenon for transporting electric and magnetic energy from one point to another. The physics of electric and magnetic fields is summarized by Maxwell’s equations, which predict the existence of the energy transport by means of waves. To every point in space one can assign a value to electric and/or magnetic field strength. The waves are simply spatial and temporal variations in the field that propagate at the speed of light.

The general solution of the wave equation provides the values of the electric and magnetic field vectors in space and time. The two vectors constitute a tensor for each point in space. Fortunately, for the discussion here, we can, by certain specialization and approximation, actually describe light with simple scalar quantities. The most elementary solutions to the wave equation are sinusoids, and these describe laser light with sufficient accuracy to be tractable to many practical cases. When a coherent light wave passes through a region in space, the electric and magnetic fields vary sinusoidally with time at any point in the region, and they vary sinusoidally along any straight line through the region.

Wavefronts are defined as surfaces of equal phase in the region where the electromagnetic field exists. The surfaces are normal to the direction of travel for the waves of interest here (transverse waves). The wavefronts are separated by one wavelength and they travel at the velocity of light. When the wavefronts are planes, the light is said to be collimated and when they are spherical, the light is said to be converging or diverging. If the wavefronts are other than plane or spherical, they must be described mathematically with a series of sinusoids.

1.1.2 Properties

Light can be described by vectors and scalars that can be manipulated and used in diagnostics.

1.1.2.1 Polarization - For the cases to be considered here, the electric and magnetic vectors are perpendicular to each other, they are perpendicular to the direction of travel of the wave, and their magnitudes are proportional to each other (transverse waves). This allows description of the wave in terms of either vector. Even with these constraints, the vectors can still rotate about the direction of travel. The state of polarization of a wave describes the direction of the electric field vector. Linearly, polarized light has a single, fixed electric field direction. Elliptically, polarized light has a rotating electric field vector. Unpolarized light is a blend of many
possible polarization states. This property of light is important because it affects the way two light waves interfere with each other as well as the way light reflects from and transmits through materials.

1.1.2.2 Amplitude and Intensity - At any point in space, the amplitudes of the electric and magnetic fields in an electromagnetic wave vary at an extremely high frequency (typically $10^{14}$ to $10^{15}$ Hertz). A linearly polarized, plane wave that has a single wavelength, and which travels in a direction $\mathbf{E}$ parallel to the $x-z$ plane in a uniform medium, can be described mathematically as

$$u(x,z,t) = a_0 \exp[i(nkx - \omega t + \theta_0)]$$  \hspace{1cm} (1.1)

where $\theta_0$ is the phase of the wave at the origin at $t = 0$; $k$ is the vacuum wavelength; $a_0$ is the wave amplitude maximum; and $n$, the refractive index of the medium, is the ratio of the velocity of light in a vacuum to that in the medium (see Figure 1.1). Equation 1.1 is an elementary solution (or eigenfunction) of Maxwell's equations, and describes a wave which is said to be coherent with itself and with all other waves that have the same wavelength.

![Figure 1.1 Propagation of a plane electromagnetic wave. The wave propagates in a direction normal to the wavefronts in the direction of $\mathbf{E}$.](image)

In principle, one can measure the electric or magnetic field in the wave. However, the variation is much too fast for ordinary instrumentation. The energy flux in such a wave is proportional to the square of the electric field, and, therefore, it varies at the same frequency. A quantity which is measurable by normal sensors or photographic material is the time averaged energy flux, which will hereafter be called the "intensity" of the wave. When averaged over times which are long compared to the period of the wave, it is given by

$$I(x,z,t) = uu^*$$  \hspace{1cm} (1.2)

This equation will be used many times in the following chapters to determine the observable in an instrument. In each case, the complex amplitude at a point is determined by vector addition of the amplitudes of various waves contributing to the total field at the point. Then the observable is computed by using Equation 1.2.

1.1.2.3 Coherence - When two electromagnetic waves are mixed, their electric and magnetic vectors add. If the waves have the same wavelength, then the difference in phase of the two waves will be constant at any location in space. When this is true, the intensity of the wave is maximum when the two waves are in phase, and minimum when they are 180 degrees out of phase. Also the intensity is maximized when the polarization direction of the two waves are the same. If the two polarization vectors are normal to each other, then interference does not occur and the intensity is simply the sum of the intensities of the two individual waves.

**EXAMPLE:** The maximum net intensity, in a region where two equal amplitude, coherent, polarization aligned waves are mixed, is four times the intensity in one of the waves (in phase case), while the minimum intensity is zero (180 degrees out of phase).
before the two waves come back to the same relative phase, it is easy to compute a beat frequency

\[ f = \frac{C}{a_1(a_2)} \]  

(1.3)

**Figure 1.2** Appearance of an electromagnetic wave comprising several frequency components.

If such a wave is split into two equal intensity parts, and if the two parts are remixed after having traveled different distances, then the resulting intensity depends upon the relative phase of the beat waves. If the beat waves are 180 degrees out of phase, then the resulting interference is effectively reduced since the amplitudes of the two waves differ at all times, making it impossible for either of the two waves to cancel the other. As more frequencies are introduced to construct the wave, the distance over which the two waves can efficiently interfere becomes smaller and smaller. The distance of efficient interference is called the coherence length of the wave. If a wave is split into two waves, and if the two waves are transmitted over different distances before being remixed, the interference (fringes) will be observed to deteriorate as the difference in distance increases. That distance difference, at which interference ceases to be observable, is called the temporal coherence length of the light wave.

A similar effect occurs when one attempts to interfere waves extracted from different places on the wavefront; that is, from two places taken from a plane which is normal to the direction of travel. As the distance is increased, the interference deteriorates until it is no longer observable. This distance is called the spatial coherence interval.

### 1.2 Interference

#### 1.2.1 Plane Waves

We now examine the interference of two coherent plane waves. Cast Equation 1.1 into Cartesian coordinates and examine the effects in the xy plane, which is normal to z, the direction of travel. In the following, we carry only the z and t coordinates since dependence on x and y is the same.

\[ u(x,z,t) = u_0(x,z) \exp[i(kn(x \cos \alpha + x \sin \alpha) - \omega t + \phi_0)] \]  

(1.4)

The intensity is proportional to

\[ I(x,z,t) = uu^* = |u_0(x,z,t)|^2 \]  

(1.5)

We will consider cases in which I is a slowly varying function of time, where I is time averaged over times large compared to the period of the wave but small compared to measurable time. Consider the mixing of two waves of like polarization traveling along
two different directions. The rigorous solution for the resulting intensity is not simple, but an approximate solution for small angles between the two directions of travel can be found simply by adding the complex amplitudes and squaring.

\[ I(x, y, t) = (u_1 + u_2)(u_1^* + u_2^*) = |u_{01}|^2 + |u_{02}|^2 + u_1 u_2^* + u_1^* u_2 \]  

(1.6)

Combining Equation 1.4 and 1.6 leads to

\[ I(x, y, t) = u_{01}^2 + u_{02}^2 + 2u_{01}u_{02} \cos \left[ \pi n(k_1 \cos \alpha - k_2 \cos \beta) \right] + \pi n(k_1 \sin \alpha - k_2 \sin \beta) \]  

(1.7)

Equation 1.7 describes a set of alternate bright and dark fringes that are not stationary in time. We shall now consider some special cases and useful examples of Equation 1.7.

**Case 1 - Same Wavelengths and Phase at the Origin**

\[ k_1 = k_2 = k, \ \Delta \phi = \phi = 0 \]  

\[ I(x, y) = u_{01}^2 + u_{02}^2 + 2u_{01}u_{02} \cos \left[ \pi n(k_1 \cos \alpha - k_2 \cos \beta) \right] + \pi n(k_1 \sin \alpha - k_2 \sin \beta) \]  

(1.8)

The resulting set of fringes is illustrated in Figure 1.3. They comprise sheets of light and dark in space. The equation for these planes of equal intensity is found by setting the argument in Equation 1.8 to a constant. Thus, for \( n = 1 \),

\[ x = \frac{(\cos \alpha_1 - \cos \alpha_2)}{(\sin \alpha_1 - \sin \alpha_2)} z + \text{const.} \]  

(1.9)

![Interference Planes](image)

**Figure 1.3** Interference of two plane waves. Equally spaced fringe planes bisect the angle between the two beams. The fringes are stationary for two beams of the same wavelength. But at different wavelengths fringes move normally to the planes.

For two particular planes the fringes are especially simple.

(a) One Beam Normal to the Plane of Observation (\( \alpha_1 = 0 \))

\[ I(x, y) = u_{01}^2 + u_{02}^2 + 2u_{01}u_{02} \cos \left[ \pi \left( 1 - \cos \beta_2 \right) - \pi \sin \beta_2 \right] \]  

(1.10)
EXAMPLE: In any xz plane of observation the fringe frequency is

\[ f_x = \frac{\sin a}{\lambda} \]  

(1.11)

At \( a = 0.5 \) microns, the fringe frequency for \( a = 10 \) arc seconds is approximately 1 fringe/cm, while at \( a = 10 \) degrees, \( f \) is 3,600 fringes/cm.

The fringe planes are given by

\[ x = \tan \left( \frac{a_x}{2} \right) \text{ + const.} \]  

(1.12)

Therefore, they are parallel to the bisector of the angle between the beams. (This is true regardless of the observation plane.) Note that when \( a = 180^\circ \) (the beams are mixed from opposite directions), the fringe planes are perpendicular to the directions of travel of the beams. The fringes are separated by 1/2.

(b) The Plane of Observation is Normal to the Bisector of the Angle Between the Two Beams (\( a_1 = -a_2 \))

\[ I(x) = u_0^2 + u_0^2 + 2u_0u_2 \cos \left[ 2nk(x \sin a) \right] \]  

(1.13)

The fringe planes are given by \( x = \text{constant} \) and are, therefore, parallel to the z-y plane. The fringe frequency, just twice that given in the previous case, is \( (2n \sin a)/\lambda \).

Case 2 - Same Phase at the Origin, Approximately Same Wavelengths, \( a_1 = -a_2 \)

\[ I(x, z, t) = u_0^2 + u_0^2 + 2u_0u_2 \cos \left[ 2nk \sin a + 2n \omega t \right] \]  

(1.14)

Two important features are to be noted. First, the lack of z dependence places the fringe planes parallel to zy plane as in Case 1b. Secondly, the argument of the cosine term in Equation 1.14 will be constant when

\[ x = \frac{2\Delta \sin a}{2k \sin a} \]  

(1.15)

which can be interpreted as a motion of the fringe planes perpendicular to themselves (x direction) with velocity

\[ v = \frac{\Delta \sin a}{2k} \]  

(1.16)

Therefore, if a tiny, stationary scatter center is being illuminated by two such waves, the scattered light intensity will oscillate with frequency \( f \), where \( f \) is the frequency difference of the two waves. If the particle moves in the direction of fringe motion, \( f \) will decrease. Motion in the opposite direction increases \( f \). This principle is the basis for determining the particle velocity and direction.

Case 3 - The Two Waves Travel Along the Same Line

\[ I(x, z, t) = u_0^2 + u_0^2 + 2u_0u_2 \cos \left[ 2n(k_1 - k_2) - 2n \omega t \right] \]  

(1.17)

At any position, z, the intensity varies at a frequency equal to the difference frequency of the two waves. This describes the wave illustrated in Figure 1.2.

If the two waves have the same wavelength, then Equation 1.17 reduces to

\[ I(x, z, t) = u_0^2 + u_0^2 + 2u_0u_2 \cos \left[ 2n \omega t \right] \]  

(1.18)

and the intensity depends upon the phase difference of the two waves. Most interferometers can be explained in terms of the above equations.

1.2.2 Spherical Waves

The time independent amplitude in a spherical wave source located at \( r = 0 \) can be expressed as

\[ u(r) = u_0 \exp (i \omega t)/r \]  

(1.19)

The quadratic approximation of this expression is

\[ u(x, z) = u_0 \exp \left[ i \frac{2n}{3z} (x - x_p)^2 \right] \]  

(1.20)

where \( x_p \) is the location of the source and \( x \) is the observation point. If two such waves are mixed, the resulting interference fringes are curved. The amplitude is
\[ u(x,z) = u_0 \exp\left[i\frac{\pi}{\lambda_1}(x-x_p)^2\right] + u_0^\prime \exp\left[i\frac{\pi}{\lambda_2}(x-x_p^\prime)^2\right] \] (1.21)

and the intensity is

\[ I(x,z) = u_0^2 + u_0^\prime 2 + 2u_0u_0^\prime \cos \frac{\pi}{\lambda_1} (x-x_p)^2 + (x-x_p^\prime)^2 \] (1.22)

Note that when a spherical wave is mixed with a plane wave, the intensity in the plane of observation is

\[ I(x) = u_0^2 + u_0^\prime 2 + 2u_0u_0^\prime \cos \frac{x(x-x_p)^2}{\frac{\lambda_1}{2}} \] (1.23)

1.2.3 The Moiré Analogy

A useful and instructive moiré analogy can be constructed with grid patterns. These have proven so useful in instruction that a set of grids has been included in the appendix to allow the reader to construct his own set of transparencies. Representing a plane wave with a set of equally spaced lines in which the direction of travel is normal to the lines (as in Figure A1), one can then produce a transparency with this as a master. When the transparency is superimposed on the master, a set of moiré fringes can be observed as shown in Figure 1.4. Note that the fringes bisect the angle between the directions of travel of the two waves as predicted. As the angle between the two directions is increased, the fringe density increases. If the relative phase of the two waves is shifted by shifting position of the master, the fringes shift as predicted. Many of the intricacies of interference can be illustrated by using the grids and a small amount of ingenuity.

![Figure 1.4 Mixing of two waves at an angle. This moiré pattern represents fringes seen from the top view of Figure 1.3.](image)

Figure A2 is included to represent a wave of slightly different frequency. By superimposing a transparency of A2 onto A1, the beat waves described above are seen as shown in Figure 1.5.

![Figure 1.5 The beating of two waves of slightly different frequency. One of the wavefronts continues to advance ahead of the other moving in and out of phase each time it advances by one-half wave.](image)
Figures A3 and A6 are included to represent spherical waves. Now superimpose a transparency of A4 onto A4. The resulting fringes are illustrated in Figure 1.6. Note how the fringes move, as a relative motion between two light sources is simulated by shifting the transparency around. The mixing of two spherical waves of different frequency causing a beat wave is illustrated by superimposing Figures A3 and A4 as shown in Figure 1.7. We will return to the analogy later to explain holography and laser Doppler anemometry.

![Figure 1.6](image)

**Figure 1.6** The interference of spherical waves of the same wavelength.

(a) Two waves of different wavelength emanating from separate points.  
(b) Two waves of different frequency emanating from the same point.

![Figure 1.7](image)

**Figure 1.7** Interference of two spherical waves of different wavelength.

1.3 Interaction of Light with Matter and Boundaries

1.3.1 Diffraction

When light interacts with matter or with a boundary, the resulting field becomes much more complicated. This constitutes the placing of boundary conditions on the Maxwell equations. The solutions then become series expressions or integrals instead of simple sinusoids. It is useful to think of diffraction as an interference phenomenon. When light passes over a boundary such as an edge, the light which strikes the edge is
scattered in every direction and this light mixes with the unscattered light. So any given point in a plane of observation receives light of two origins, scattered and not scattered. Since the distance to the point is different for the two sources, the two components of light will have different phases and can interfere constructively or destructively, depending on the phase difference. Therefore, one expects to observe fringes in the vicinity of an edge (Figure 1.8a). There is no such thing as a "shadow" in the usual sense for coherent light, only a diffraction pattern.

Consider the appearance of light after having passed through an aperture of diameter, d (Figure 1.8b,c). Near the aperture, the light distribution is fairly uniform across the aperture. As one moves further along the optical axis, a series of diffraction rings can be observed. This is called the Fresnel region, or the near field, and the pattern will continue to exhibit this appearance as long as the distance is less than about $d^2/4$, a distance which is appropriately termed one far field distance or the distance to the far field. When the observation distance reaches and passes the far field distance, the appearance of the diffraction pattern changes drastically. The light which was scattered from the edge remains in phase with the unscattered light over a much larger region of the observation plane. The intensity is given approximately by

$$I(r) = \frac{2 j}{(4\pi r)} \left( -\frac{\lambda z}{r^2} \right)$$

(1.24)

A central part of the diffraction pattern (sometimes loosely referred to as Airy's pattern), containing about 84% of the energy, dominates and, for a fairly large distance, is actually smaller than the aperture itself. Its radius increases with z and is given by

$$r_m = 1.22 \frac{\lambda z}{d}$$

(1.25)

so that the half divergence angle, $\theta_d$, of this portion of the beam is given by

$$\tan \theta_d = \frac{\lambda z}{d}$$

(1.26)

A similar result can be observed when light passes over an edge or an obstructing object. Light scattered from an edge has a phase difference, $\Delta \phi$, with unscattered light. For the small angle case, this difference along the z-axis is given by

$$\Delta \phi(z) = \frac{2 \pi x^2}{\lambda z}$$

(1.27)

Figure 1.8a Interaction of light with edges. A diffraction pattern can be observed on a screen located at z. Light scattered from the edge (top figure) adds in phase with unscattered light when it travels a distance $n\lambda$ further producing a bright fringe. Light scattered along the centerline axis (bottom figure) produces a bright center when it is in phase, having traveled $n\lambda$ further than unscattered light. After the distance $z$, where the path difference of scattered and unscattered light is much less than 1, the center stays bright. This is called the far field.
1.3.2 Reflection and Refraction

When light strikes a surface, some of the energy may be reflected from the surface, some transmitted into it, and some may be absorbed by it. If the surface is a perfect dielectric material, no absorption takes place as long as the intensity of the field is not strong enough to tear away the bound electrons. Figure 1.9 summarizes the interaction of the light with a dielectric surface. The angle of incidence and the angle of reflection are equal. The angle of refraction, \( \theta \), is given by Snell's law:
Figure 1.9 Reflection and refraction of light at a dielectric surface. Reflection and transmission depend upon polarization direction.

The incident ray, transmitted ray, and reflected ray stay in a single plane, the "plane of incidence." If the light is divided into its polarization directions, one component in the plane of incidence and one normal to the plane of incidence, then the two components can be treated independently. For example, if all of the light is polarized with a polarization vector in the plane of incidence, the polarization of the transmitted and reflected light remains in this plane.

A few simple special cases occur so often that it is useful to mention them here. There is an angle of incidence, called Brewster's angle, at which light polarized in the plane of incidence is totally transmitted (Figure 1.10). This angle is given by

$$\theta_B = \tan^{-1} \frac{n'}{n}$$

(1.30)

Figure 1.10 Refraction at the Brewster angle. Light polarized in the plane of incidence is transmitted completely.

When light travels toward a material of lower refractive index, there is an angle called the angle of total internal reflection, at which all light is reflected (Figure 1.11). This angle is given by

$$\theta_c = \frac{n'}{n}$$

(1.31)

These two conditions are used extensively in experimental optics.

1.3.3 Propagation

Because of diffraction, a light wave changes in intensity as it propagates through space. The case of the uniform circular light wave has already been treated above, and the intensity distribution at some propagation distance, \( z \), was given by Equation 1.2A. Another case of practical importance is the case of a light wave which has a Gaussian intensity distribution. This particular distribution has the unique and useful property that it maintains its form while propagating through space. The intensity has the form
Figure 1.11 Total internal reflection. Beyond a critical angle light traveling from a higher to a lower refractive index is totally reflected.

\[ I(r) = I_0 \exp\left(-2r^2/a^2\right) \]  

(1.32)

where \( I_0 \) is the central intensity and \( a \) is the radius to the \( 1/e^2 \) intensity. The radius increases according to the relation

\[ a(z) = a\left[1 + \left(\frac{\lambda}{\sin^2 \theta_c}\right)^2\right]^{1/2} \]  

(1.33)

The propagation direction of a light wave will change if the refractive index of the medium in which the wave is traveling changes along the wavefront. This can be explained by recalling that the direction of travel is normal to the wavefront and that the wavelength becomes shorter in a medium of higher refractive index (Figure 1.12). The light wave is deviated by an angle

\[ \frac{dz}{dz} = \frac{1}{dn} \frac{dn}{dz} \]  

(1.34)

The second integral is called the optical pathlength from \( 0 \) to \( z \). It is the number of wavelengths existing between \( 0 \) and \( z \) times the vacuum wavelength. For example, since the wavelength is shortened in a medium with refractive index greater than unity, the optical pathlength becomes larger. In vacuum, optical pathlength is, of course, just pathlength. In many materials, \( n \) can be related to density, \( \rho \), by the Dale-Gladstone constant, \( G \).

\[ n = 1 + G \rho \]  

(1.36)

so that
These changes alter the amplitude and the phase of the light over an observation plane. The use of this effect to observe flow fields will be described in Chapters 2 and 3.

1.3.4 Speckle

When a light diffusing surface is viewed or photographed in incoherent light, it exhibits a more or less smooth appearance. When the same surface is viewed or photographed in coherent light, it exhibits a grainy appearance. This graininess is a statistical phenomenon that is called speckle. There are two categories of speckle, one caused by the limited resolution of the viewing system and the other caused by the limited extent of the illuminating system (Figure 1.13). When a finite sized diffuse surface is used to illuminate another surface that is located at a distance, \( z \), the intensity on the second surface will exhibit speckle having a mean diameter of

\[
d_{sp} = 1.22 \frac{\lambda z}{D}
\]

The phenomenon can be explained in the following way: any point on the observation plane is illuminated by a confluence of light that originated at the many different points on the diffusor. Each of these points is characterized by a phase that is different from other points in a random way. There is a probability that the sum of the sources will add up to produce a bright point in the observation plane; and, likewise, there is a probability that a dark point will be produced.

When a diffuse surface is imaged, every resolution element in the image receives light from a large number of randomly phased sources that make up the resolution points (such as the high and low points that constitute the surface texture). There is a probability that the different components of light making up the resolution points will either add up sufficiently in phase to produce a bright point or out of phase to produce a dark point. The average speckle size, therefore, can be expected to be the size of the resolution of the imaging system.

\[
d_{sp} = 1.22 \frac{\lambda}{d_t}
\]

The resulting useful resolution of the imaging system is, therefore, actually three or four times less than that given by the normal equation for resolution since the image will consist of speckles of that size. The highest useful resolution is attainable when direct, not diffuse, illumination can be used so that speckle does not occur.
1.4 Control of Light

In this section, we shall summarize and discuss some of the most basic elements for the control and use of laser light. The special characteristics of laser light provide a spatial and temporal controllability that cannot be attained with incoherent light and have fostered many new applications. In general, coherent light can be delivered into smaller spatial regions, at higher intensities, in smaller time elements, over greater distances, with more precise modulation, and, because of single wavelength, can be separated more easily from background radiation. The following discussion covers some of the devices and physics that make these important characteristics possible.

1.4.1 Lenses and Mirrors

Lenses and mirrors transform an input light wave from one form, such as planar, to another such as converging. The input space is called the object space, while the output space is the image space. The object space can be emulated exactly in image space only for very special cases; however, the approximations are useful especially for engineering design.

The most basic function of a lens is the conversion of a plane wave to a converging wave. When a plane wave strikes a positive lens, it converges to a focal point that is located at a distance from the lens, f, called the focal length of the lens. When the wave strikes a negative lens, it appears to diverge from a point on the input side of the lens, also called the focal point.

Points in the object space are approximately related to points in the image space by the thin lens equation.

\[ s' = \frac{sf}{s - f} \] (1.41)

Where \( s \) is the distance from the object point to the lens, \( a' \) is the distance from the lens to the image point, and \( f \) is the focal length of the lens.

The image is never exactly the same as the object. Three things limit the accuracy with which a wavefront can be changed from one form to a second desired form: 1) the design of the lens (such as the specification of material, surface curvatures, and the combinations of these), 2) the accuracy of manufacture of the lens, and 3) diffraction. The effects introduced by the first two are called geometrical aberrations, and the limits set by the third are called diffraction limits.

A high quality, well designed lens, in which geometrical aberrations are small compared to the limiting effects of diffraction, is loosely termed a diffraction limited lens. A lens cannot be produced to be diffraction limited for all transformations. The most common diffraction limited lens is a collimator; it transforms a plane wave into a converging wave or vice versa. A parabolic mirror is a diffraction limited collimator. Lenses and mirrors must be illuminated from the correct direction to produce diffraction limited images. Additionally, lenses must be illuminated from the correct side.

A converging wave passes through a transition region of minimum beam diameter wherein the wave is plane before beginning to diverge. The intensity distribution at the focus depends upon the intensity distribution across the lens. Figure 1.14 a-d summarizes the results of illuminating a lens with two different plane wave intensity distributions, uniform and Gaussian. Note that the effective focus diameter and position depends upon the diameter of the input beam. The lens performs a focusing action on the beam while diffraction causes a diverging action. The net result is that the minimum occurs at a different point than would be predicted from geometrical optics alone. Since many lasers produce a wavefront of Gaussian intensity profile, it is extremely important to understand Gaussian optics and the difference between geometrical optics predictions and the more accurate diffraction optics results. Geometrical optics assumes that light can be focused to an infinitely small point. This, of course, does not happen. The image distance in diffraction theory actually depends on the input beam diameter and intensity distribution. The following are some significant characteristics of beams with Gaussian intensity profile.

1. Propagation through empty space - the intensity profile remains Gaussian, a unique property of Gaussian intensity profiles.
2. Divergence from a source located at the front focal plane - the effective image position (minimum diameter) is at the back focal plane.
3. Truncation - the beam will not remain Gaussian if it is truncated. The Gaussian should not be truncated below two \( \frac{1}{e^2} \) diameters to retain Gaussian properties.
4. Energy removal by truncation - a curious result is that the percentage of total energy removed by truncation is the same as the percentage of maximum intensity at which truncation is done.
Figure 1.14 Focusing laser beams.
Example: A Gaussian beam passes through a circular stop. The intensity of the beam at the edge of the stop is one percent that at the center. Therefore, one percent of the energy will be removed from the beam.

At the focal point, 84 percent of the energy passes through the circle with diameter $d_A$ for uniform illumination, and for a Gaussian profile 87 percent passes through the circle with diameter $d(e^{-2})$. The distance in the focal region over which the waves are nearly plane is referred to as the depth of field $\Delta f_e$. For the Gaussian case,

$$\Delta f_e = \frac{\lambda f_s}{\pi a^2}$$

(1.42)

1.4.2 Scanners, Deflectors, Modulators, and Shutters

Laser beams can be scanned, deflected, and modulated at high speeds with mechanical and electro-optic devices. Mechanical devices include rotating or vibrating prisms, gratings, and mirrors that can scan a beam over wide angles at kilohertz rates. Vibrating mirrors are attached to electromagnetically driven components that oscillate at rates up to tens of kilohertz. The highest rates are attained when torsion elements are driven at resonance, but it is possible to drive these at programmable angles at kilohertz rates. Rotating elements can attain even higher rates. Holographic gratings or other holographic optical elements also have provided an extremely useful range of capabilities.

Electro-optical scanners can deflect beams at megahertz rates. The most common of these is the Bragg cell, an acousto-optical device. In this device, the beam is diffracted from ultrasonic waves that are generated in a material by piezoelectrically driven crystals.

Other devices modulate the polarization properties of light at megahertz rates. When used in conjunction with polarization filters or polarization dependent reflectors, the beam can be deflected or shuttered at the same rates as the polarization modulation.

Other types of modulators modulate phase only. A Pockels cell can be used in this mode. Any element that changes optical pathlength can be used as a phase modulator. A somewhat unique phase modulator is an optical fiber that is stretched electromechanically or piezoelectrically while light passes through it. By wrapping the fiber around the
piezo-electric element, a full wavelength shift in the optical pathlength of light traveling through the fiber is attainable with a small amount of voltage applied to the crystal.

Coherent light can be shuttered more easily than incoherent light, because more parameters are controllable for the shuttering process. Shuttering elements employ both mechanical and electro-optical techniques covering a vast range of shuttering times down to nanoseconds or less. One simple form of shutter employs a scanner, as described above, to sweep a focused beam across a small aperture. Shuttering times less than one microsecond are possible with this method. Such times are attainable even with mechanical shutters if the beam is focused to small enough size. When the size of the beam cannot be made small, mechanical methods are limited to the millisecond regime.

The fastest shutters are electro-optical devices. A Pockels cell, when used as described above, can shutter a laser beam for times as short as nanoseconds.

Pulsed lasers themselves can provide the shuttering action if all other light is blocked from the sensor. This can be accomplished by using narrow band filters that pass only the laser wavelength and a narrow band around it. The effect can be further enhanced if the room light is limited to a band that is widely different from the laser. Common fluorescent lights emit little or no radiation at wavelengths above 650 nm, with the result that a high pass cutoff filter such as Scott RG650 can be used to cover the film or sensor in a fully lit room when a laser above this wavelength is used.

1.4.3 Holographic Optical Elements

If any two wavefronts of light are known, then a hologram can be produced that converts one of the wavefronts into the other. Holographic optical elements (HOE) are holograms that are designed to convert a wavefront of one specific type into a different desired wavefront. A HOE is made by using one of the two wavefronts as a reference wave and the other as the object wave. Then, when the hologram is illuminated with the reference wave, the desired wavefront emerges.

HOEs can be used to perform the function of virtually all optical elements such as gratings, lenses, and mirrors, or even much more complex devices. Primary advantages include low cost, small thickness and weight, and the fact that they can be produced on any shape of surface. The major disadvantage is that they are usually not achromatic.

1.4.4 Spatial Filtering

Laser beams almost always contain diffraction noise, flashlamp or plasma illumination, and transverse mode structure. Much of this can be removed by a procedure of spatial filtering. The beam is focused by a high quality lens through a tiny pinhole (Figure 1.15). The intensity distribution of the focused laser beam depends on the wavelength, lens diameter, focal length, and intensity distribution across the lens.

![Figure 1.15 Spatial filtering](image)

Figure 1.15 Spatial filtering. Rays that are not collimated in the input beam are removed by the pinhole.

The pinhole diameter can be estimated by examining Figure 1.14. The aperture having the diameter of the Airy disk will pass 84 percent of the collimated light striking the lens, while removing most of the uncollimated light. Actually, most laser beams of interest are Gaussian in intensity profile rather than uniform. For continuous wave (CW) lasers, pinhole diameters range from a few micrometers in diameter upward, depending upon f. For high power pulsed lasers, the lens focal length must be increased to reduce the intensity at focus, requiring that the pinhole size be increased. Otherwise, the intensity can become high enough to ionize the air or damage the pinhole. Air breakdown does not necessarily render the process useless. The plasma absorbs the light and shortens the pulse length and may eventually destroy the pinhole, but the process works at least until the pinhole is destroyed. The use of diamond pinholes lessens the problem.
Figure 1.16 illustrates a spatially filtered beam with various size pinholes. To emphasize the filtering action, a 10 micrometer diameter wire was stretched across the back side of the 20 power microscope objective. With a 30 micron diameter pinhole, almost all of the diffraction noise is removed. With no spatial filtering, not only is diffraction noise present in the beam, but also multiple reflection interference rings are seen.

![Diagram of microscope setup with pinholes and wires.](image)

Figure 1.16 Spatial filtered beams with various pinhole sizes. Note that higher spatial frequency information passes into the final image as the pinhole gets larger.

1.4.5 Illuminating an Object Field

Three different classes of coherent illumination, direct, diffuse, and structured can be used. Diffuse illumination suffers the problem of speckle, but provides the easiest viewing by eye. Direct illumination provides speckle free light, but is more difficult to view by eye. The most common form of structured light is the so-called sheet of light or light screen produced with a cylindrical lens. This provides a convenient way to view activity in a plane, without suffering the optical noise of light scattered from the remainder of the volume that is not under examination. Other forms of structured light include multiple sheets of light or rays of light.

Other choices of object illumination include front lighting and backlighting. Backlighting provides for high resolution and high contrast viewing of the perimeter of an opaque object, while front lighting provides for the viewing of surface detail of the object.
1.5 Lasers

1.5.1 How They Work

Lasers are characterized by two major features, the lasing medium and the physical arrangement that forms a cavity, or interferometer, for selecting photons of a particular frequency and feeding them back into the medium, where they are amplified before leaving the cavity. The technical knowledge required to build a laser became available with Einstein's theory of stimulated emission in the early 1930s. Nevertheless, the first operational laser was built by Maimon in 1960, nearly a half century later. Within months of the announcement of his achievement of lasing action in ruby, investigators from laboratories the world over reported lasing in a large number of other materials. Today, hundreds of commercially available lasers employ many different lasing materials to provide useful characteristics. In the following paragraphs, the individual phenomenological elements of a laser are discussed and then combined to produce a laser.

1.5.1.1 Stimulated Emission - An excited molecule can decay to its ground state by any one of several different mechanisms. The decay must be accompanied by an energy transfer of the residual energy, \( \Delta E \), either to another molecule (radiationless transition) or to an emitted photon. The emission of a photon can occur in either a spontaneous process or a stimulated process. The probability for stimulated emission at radiation frequency, \( f = \frac{E}{h} \), where \( h \) is Planck's constant, is increased sharply when the excited molecule is illuminated by radiation very close to \( f \). These photons are said to be resonant with the excited state, and their presence greatly shortens the average lifetime of the excited state. Stimulated radiation leaves the excited molecule in phase with the stimulus field (Figure 1.17). Transitions are possible only between selected energy levels. When a molecule is struck by a resonant photon, the probability for absorption of the photon by an excited molecule is the same as the probability for emission of the same photon by an excited molecule.

![Molecular excitation and decay](image.png)

Figure 1.17 Molecular excitation and decay.

1.5.1.2 Population Inversion - The lasing medium can be excited in many ways, but the resulting excited material is characterized by the population of molecules that excitons through an excited material, some of them will be absorbed and some of them will stimulate emission of more of the same photons, resulting in an amplification of the illuminating radiation. The condition required for a net amplification is simply that the number of stimulated emissions exceed the number of absorptions and other losses. Since the probability of stimulated emission and absorption is the same, this condition occurs when the number of molecules that populate the excited resonant higher state exceeds the number of molecules that populate the lower state. This condition is appropriately described as population inversion between the two states, since the condition is the inverse of the usual stable condition.

1.5.1.3 Amplification - When resonant photons pass through a material where population inversion exists, the net energy in the traversing beam will increase. The material acts as a light amplifier for photons that have energy equal to the difference of energy between the population inverted states. Amplifiers are used to increase the output power of a laser as well as to aid in the amplification and detection of weak coherent light waves. In some instances a laser may incorporate three or four amplifiers in tandem.

1.5.1.4 Feedback-The Oscillator - When the lasing material is placed between two reflectors, the amplified light wave is fed back into the material where it is further amplified. If a light wave gains in strength after making a round trip between the two mirrors, the lasing threshold is said to have been exceeded and lasing action takes place. The two mirrors may be planar, curved, or combinations of plane and curve.

The space between the two mirrors is called the cavity. The energy in the cavity could become so high that it destroys the mirrors or the lasing material, but by making one or both of the mirrors partially reflecting, some of the light leaks out of the cavity (see Figure 1.18). Waves travel in both directions in the cavity, forming a standing wave whose wavelength satisfies \( L = \frac{\lambda}{2} \) where \( \lambda \) is an integer.
**ABSORPTION AND AMPLIFICATION**

**AMPLIFYING MEDIUM** 

**INPUT WAVE** 

\[ I_0(0) \]

**OUTPUT WAVE** 

\[ I_0(x) \]

**ABSORBING MEDIUM** 

**INPUT WAVE** 

\[ I_1(0) \]

**OUTPUT WAVE** 

\[ I_1(x) \]

**LEGEND:**
- \( \circ \) Atom in upper state 2
- \( \bullet \) Atom in lower state 1

\[ I_0(x) = I_0(0)e^{-ix} \]

\[ \eta(v) = (N_2 - N_1)e^{-2\text{abs}^2\text{Prop}} \]

(a) Absorption and amplification.

N PHOTONS

[Diagram of an amplifier and an oscillator with feedback]

(b) An amplifier and an oscillator.

(c) The laser cavity

Figure 1.18 Lasing Action.
1.5.2 Coherence Improvement and Control

1.5.2.1 Frequency Broadening - Temporal coherence is tantamount to single frequency; spatial coherence is tantamount to simple, plane or spherical wavefront. Improving the coherence of a laser amounts to making the cavity strongly supportive of a single frequency, simple wavefront and strongly unsupportive of anything else. If the energy associated with the inverted state were the same for all molecules, then a single frequency would result. Since the energy is not the same for all molecules, the frequency of light emitted from an excited molecule upon decay, $\Delta E/\hbar$, is not the same for all molecules. The energy broadening mechanisms include:

(a) Natural Broadening due to the uncertainty principle, which states that the product of the state lifetime and the energy uncertainty must exceed Planck's constant, giving rise to a frequency broadening of typically about 15 MHz.

(b) Pressure or collision broadening is the alteration of the energy of a molecular state, through collisions or close proximity with other molecules. This effect is most important in solid laser materials or gas lasers having higher pressures (>1 torr).

(c) Radiation Broadening related to the finite width of the stimulated emission resonance curves.

(d) Zeeman Broadening caused by magnetic fields which alter the energy states.

(e) Doppler Broadening caused by the different kinetic energies of the molecules that are transferred to the emitted photon, typically of the order of several GHz and greater for higher temperature lasers.

1.5.2.2 Transverse Electromagnetic Modes (T.E.M.) - The solution of the electromagnetic cavity equations is a series of eigenfunctions with indices $p, l, q$. Each of these functions represents a mode of operation of the laser and is called a transverse electromagnetic mode (T.E.M.). A laser may operate in many modes at once, since the general solution of the equation is a combination of the eigenfunctions; however, the lower order modes usually predominate. A laser can be forced to operate predominately in a single mode, usually $p = 0, l = 0, q$, or the TEM$_{00}$ mode. The various $p$ and $l$ modes are characterized by different intensity distributions about the optical axis of the laser. The amplitude of the $p$ and $l$ mode for a laser with cylindrical geometry is given by

$$U(r, \theta) = U_0 \left( \frac{r}{a} \right)^{l-1} L_p \left( \frac{2a^2}{\sigma^2} \right) \exp \left( -\frac{r^2}{\sigma^2} \right) \sin^l \theta,$$

$$U_0 \quad \text{amplitude at the optical axis}$$

$$L_p \quad \text{Laguerre Polynomial}$$

$$\sigma \quad \text{radius}$$

$$\theta \quad \text{angular coordinate}$$

Consider the $p = 0, l = 0$ case. Since $t^0(x) = 1$

$$U_0(r, \theta) = U_0 \exp \left( -\frac{r^2}{a^2} \right).$$

Therefore, the lowest order mode has a Gaussian amplitude and intensity profile.

The frequencies of the various $t$, $p$, and $q$ modes differ, the amount depending upon the laser geometry. If, for example, the laser mirrors are plane, the frequency does not depend upon $t$ and $p$. In general, the various modes are not coherent with each other. The lowest order mode is most widely used because the Gaussian intensity distribution is least divergent, more easily focused, and is most easily used in calculations; the higher the mode structure, the greater the divergence rate. In practice, a laser is forced to operate in the TEM$_{00}$ mode by making the cavity optics almost perfect, and by using apertures to restrict the lasing to a small diameter.

1.5.2.3 Longitudinal Modes - For each molecular energy transition, a lasing medium can provide one broadened frequency or color at which the laser can, in principle, operate. All but one color can easily be removed by using selective mirrors or prisms within the cavity. Within the broadened frequency, however, lies a number of individual frequencies that must be eliminated to approach a single frequency from the laser. These are denoted by the $q$ index on the TEM$_{plq}$. If the lasing material is placed between two mirrors that bound the cavity, standing waves are supported between the two mirrors and their nodes occur on the mirrors (Figure 1.19). The condition placed upon the standing wave is that the distance between the two mirrors be an integral number, $q$, of half wavelengths. If the medium is uniform this can be expressed as

$$L = q \lambda/2 = q c/2f$$

If $\lambda$ lies within the broadened energy transition then the laser will support that frequency.
Figure 1.19 Cavity modes and mode selectors. The standing wave supported by the cavity is an integral number of half wavelengths.

EXAMPLE: For a typical gas laser, say $\lambda = 0.6$ micrometers, then $f = 5 \times 10^{14}$ sec$^{-1}$, and if $L = 1$ meter, then $q > 3 \times 10^4$. Adjacent longitudinal modes are separated in frequency by $c/2L$ or 150 MHz in this case. If the frequency broadening for the given transition were 1.5 GHz, then one could expect to get up to 10 longitudinal modes from the laser.

Figure 1.20 illustrates this effect. The curve is the broadened frequency spectrum, and the curve is the product of the spectrum and the cavity efficiency as a function of wavelength. The width $\alpha$ is a measure of the selectivity of the two cavity mirrors known as finesse, to be discussed later. Only those modes within the area above a certain threshold can lase. The various modes are in constant competition with each other for the energy of the lasing medium and consequently adjacent modes rarely lase at the same time.

Figure 1.20 Representation of possible resonant frequencies centered about one color.
If the cavity contains a varying refractive index, then the wavelength varies within the cavity. The number of half waves between the two mirrors bounding the cavity must still be an integer. Therefore, if the refractive index varies in a direction normal to the cavity axis, or if the two mirrors are not precisely parallel, then the frequency of light likewise varies across the cavity. This is illustrated in Figure 1.21, where the cavity supports two different wavelengths on the two different halves. This is one reason why it is difficult to produce a single frequency from a large diameter cavity.

![Image](image_url)

**Figure 1.21** Spatial variation of frequency by a nonuniform lasing medium.

### 1.5.2.4 Single Frequency Lasers

- To improve the laser to its ultimate in coherence requires that all but one of the longitudinal modes be eliminated. This can be approached in one of several ways. The laser can be shortened such that the mode spacing is larger than the entire broadened spectrum. Namely,

\[
\frac{v}{n_L} > b
\]

This approach limits the cavity length and, consequently, the output power. Additionally, special attention must be given to insure that a mode does indeed lie near the peak of the envelope curve, and that it does not move away from the peak leading to instabilities or power fluctuations (to be discussed in the next section).

Other techniques for frequency selection depend upon the use of tuned reflectors (etalons) acting within the cavity or as cavity mirrors. One of the simplest is the tilted Fabry-Perot etalon which is inserted into the cavity. The transmission of an etalon has peaks at multiple wavelengths that depend upon the angle of incidence of the wavefront. These maxima are matched with one of the lasing frequencies by tilting the etalon. See Figures 1.22 and 1.23. The resonant frequency of the etalon can also be adjusted by changing the separation of the reflecting surfaces. The most common way to do this is to control the temperature of the etalon.

### 1.5.2.5 Stability and Tuning

- The latter discussion has been somewhat oversimplified, and a few qualifying considerations must be mentioned. From the previous discussion, it is clear that a small variation in \( n_L \), especially for a single mode laser, leads to a fluctuation in the power output of the laser. Maintaining the optical pathlength, \( n_L \), between the two end mirrors, exactly constant is not trivial because of thermally induced variations in \( L \) and because of fluctuations in \( n \) in the cavity.

The variations in \( n \) arise out of high frequency fluctuations in the plasma properties (for a gas laser), and little can be done to correct that. The long term stability of a laser is mostly related to the thermal and mechanical stability in \( L \). This is the primary reason why a laser requires a warm-up time to reach a constant average output power.

Tuning of an etalon changes the gain at a given frequency as is shown in Figure 1.23. If only one frequency is lasing, it will continue to lase until the efficiency for an adjacent mode overtakes it. Then the laser will "mode hop" to the new frequency. Therefore, tuning the laser cavity is a discontinuous process.

The resonant frequency of the cavity may be changed deliberately by changing the pathlength \( n_L \). The most common method is to mount one of the end mirrors on a piezo-electric crystal. The insertion of a material of time varying index of refraction into the cavity is a second method.

When a fixed single frequency of light is required, a feedback from the laser output can be used to tune the cavity to hold the output frequency constant. Several error signals are available for the feedback. One might well conclude that the laser output is at a maximum when the etalon is tuned resonant to the center of the broadening curve. Then the output of the laser could be monitored and used to drive the cavity mirror to maintain this peak. This reasoning is partially correct. The broadening curve does not accurately represent the frequency distribution when lasing takes place. Saturation effects cause a "hole burning effect" in the frequency distribution, producing a gain curve (Figure 1.24) that is characterized by a dip in the curve (termed "Lamb dip"). When the laser is operated at this point, either an increase or decrease in \( n_L \) will increase the output power of the laser. This provides an information source for feedback to drive the mirror position. The frequency can be located and fixed by this procedure to an accuracy of about one part in 10^8.
More accurate methods for location and fixing the frequency of a laser include the use of the Zeeman effect to produce a similar "dip" in the output versus frequency curve. The use of absorbing vapors and dyes in the cavity is yet another method to narrow wavelength.

1.5.2.6 Q-Switching - In the normal mode of operation, lasing action will begin as soon as a certain threshold of excitation is reached. Lasing will continue as long as the
lasing medium is "pumped". If lasing is inhibited, then the population inversion increases, so that in the excited state remaining in the lasing medium is at least one half of the molecules in the material be excited before lasing can occur. If the inhibitor is then suddenly removed, the amplification is enhanced for a short period of time, and the laser will emit a burst of light that lasts as long as is required to sweep the surplus of photons from the cavity, typically 10 to 100 nanoseconds. This switching of the cavity's Q is from one value to another is commonly used in solid state lasers to generate high power, short duration pulses of light.

Q-switches fall into two categories, light valves and beam deflectors. Beam deflectors inhibit lasing by misaligning the laser cavity or by defocusing the wavelengths in the cavity. Light valves inhibit lasing by introducing an absorber or blocking device into the cavity.

Either of the cavity mirrors can be employed as a Q-switch by tilting it out of alignment to inhibit lasing. Rotating or piezoelectrically driven mirrors are often used for this purpose. Acousto-optical beam deflectors can be used as Q-switches within the cavity.

 Electro-optical Q-switches are fast acting light valves. Two common varieties are Pockels cells and Kerr cells, which are polarization rotation devices, and, when used in conjunction with polarization analyzers, can become selective absorbers of light. The polarization rotation is achieved by the addition of a high voltage to the crystal, making it possible to turn the Q-switch on and off in fractions of a microsecond. Electro-optical Q-switches possess the distinct advantage that they can be turned on and off with extreme precision. They can be fired easily with a variability less than 50 nanoseconds. On the negative side, they are more expensive and less efficient than other types of Q-switches. Frustrated total internal reflection is yet another mechanism used for high-speed light shuttering.

Some chemicals possess absorption characteristics that make them ideal for use as Q-switches. The chemical, which must be highly absorbent at the laser wavelength, is placed in a transparent holder within the cavity. This inhibits lasing until the light absorbing solution becomes saturated (or bleached) by the impinging radiation, depleting the molecules that are available for absorption. The solution then becomes momentarily transparent to the radiation. For ruby lasers, the most common chemicals are cryptocyanine dye in alcohol or acetone solution and chlorophyll-D in mineral oil solution. The Schott Company RG-690 red cutoff dye filter can also be used as a Q-switch for ruby.

Chemical Q-switches are the least expensive of all, and, because of their narrow resonance absorption characteristics, enhance lasing at extremely narrow wavelengths. Unfortunately, they cannot be precisely controlled, have a dependence on environment and history, and are not as repeatable as other types. Their variability in time of firing is typically about 20 microseconds. Used in combination with other types of Q-switches, they can, however, enhance the coherence of a laser.

1.5.3.2 Solid State Lasers - Ruby and YAG (yttrium, aluminum, garnet) are the most common of the solid state lasers in this field. (Diode lasers are treated separately.) The dominant wavelengths are 694 nm and 1060 nm respectively, and it is possible to frequency double either to produce 347 nm and 533 nm respectively. It is further possible to triple YAG to produce 353 nm.

Because of the higher density of radiating molecules in a solid, much more energy can be produced per unit volume than in gas lasers. Solid lasers must be pumped optically by lamps or other lasers, whereas gas lasers can be pumped by passing a current through the partially ionized gas.

Ruby is termed a three level medium for lasing, because the lasing energy arises from a transition from an excited state to the ground state. Initially, the molecules are excited to a large band of energy states that decay down to a ground state, creating a population inversion with respect to the ground state. Clearly, this requires that at least one half of the molecules in the material be excited before lasing threshold can be reached. So ruby is a material that must be pumped "hard" for lasing.

Population inversion is much easier to attain between two initially unpopulated states. All that is required is that, during pumping, more excitations go to the higher state. Such materials are termed four level lasing materials. YAG is a four level material, and it has a much lower threshold than ruby.

Figure 1.25 illustrates two common ways of pumping a solid state laser. In one case, a xenon flashlamp is wrapped around the laser material; in the second case, a linear flashlamp is placed at one focus of an elliptical reflector while the laser rod is
Table 1.1
SUMMARY OF AVAILABLE LASERS

<table>
<thead>
<tr>
<th>Type</th>
<th>Medium</th>
<th>Pulsewidth (seconds)</th>
<th>Energy (Pulsed)</th>
<th>Wavelength (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Power (GW)</td>
<td></td>
</tr>
<tr>
<td>Gas</td>
<td>Helie</td>
<td>CW</td>
<td>10^-4 - 10^-1</td>
<td>633*, 1150, 3400</td>
</tr>
<tr>
<td></td>
<td>Hcd</td>
<td>CW</td>
<td>10^-4 - 10^-1</td>
<td>325*, 642</td>
</tr>
<tr>
<td></td>
<td>Argon</td>
<td>CW</td>
<td>10^-2 - 10</td>
<td>330, 488*, 514*</td>
</tr>
<tr>
<td></td>
<td>Krypton</td>
<td>CW</td>
<td>10^-2 - 1</td>
<td>330, 647*, 799</td>
</tr>
<tr>
<td></td>
<td>CO2</td>
<td>CW</td>
<td>10^-3 - 100</td>
<td>9000, 10700*</td>
</tr>
<tr>
<td>Gas</td>
<td>Argon</td>
<td>10^-6 - 10^-4</td>
<td>10^-4</td>
<td>330 - 514*</td>
</tr>
<tr>
<td></td>
<td>Cu Vapor</td>
<td>2 x 10^-8</td>
<td>.006</td>
<td>510, 578</td>
</tr>
<tr>
<td></td>
<td>Au Vapor</td>
<td>2 x 10^-8</td>
<td>.003</td>
<td>628</td>
</tr>
<tr>
<td></td>
<td>Nitrogen</td>
<td>10^-6 - 10^-9</td>
<td>up to .03</td>
<td>337</td>
</tr>
<tr>
<td></td>
<td>CO2, TRA</td>
<td>10^-7 - 1</td>
<td>up to .01</td>
<td>9000 - 11000</td>
</tr>
<tr>
<td></td>
<td>Excimer</td>
<td>10^-7 - 10^-9</td>
<td>up to 1</td>
<td>262, 308, 351</td>
</tr>
<tr>
<td>Solid</td>
<td>Ruby</td>
<td>10^-3 - 10^-8</td>
<td>10^-3 - 10</td>
<td>347, 694*</td>
</tr>
<tr>
<td>State</td>
<td>YAG</td>
<td>10^-4 - 10^-8</td>
<td>10^-3 - 10</td>
<td>265, 330, 1060*, 1318</td>
</tr>
<tr>
<td></td>
<td>Glass</td>
<td>10^-4 - 10^-8</td>
<td>10^-3 - 10</td>
<td>263, 355, 525, 532, 1060*</td>
</tr>
<tr>
<td></td>
<td>Alexandrite</td>
<td>10^-4 - 10^-8</td>
<td>10^-3 - 10</td>
<td>730 - 790</td>
</tr>
<tr>
<td></td>
<td>GSGG</td>
<td>10^-4 - 10^-8</td>
<td>10^-3 - 10</td>
<td>1060</td>
</tr>
<tr>
<td>Semi-</td>
<td>GAs</td>
<td>10^-7 - 10^-6</td>
<td>up to 10^-2</td>
<td>850 - 910</td>
</tr>
<tr>
<td>Diode</td>
<td>Lead Salt</td>
<td>10^-7 - 10^-6</td>
<td>up to 10^-6</td>
<td>3000 - 30000</td>
</tr>
<tr>
<td>Tunable</td>
<td>Ar or Kr</td>
<td>CW</td>
<td>up to 6 watts</td>
<td>250 - 1000</td>
</tr>
<tr>
<td></td>
<td>Pumped</td>
<td>YAG Pumped</td>
<td>10^-8 - 10^-3</td>
<td>200 - 4500</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Excimer Pumped</td>
<td>10^-8 - 10^-3</td>
<td>200 - 1000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Flashlamp Pumped</td>
<td>10^-7 - 10^-5</td>
<td>200 - 1000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N2 Pumped</td>
<td>10^-8 - 10^-5</td>
<td>200 - 1000</td>
</tr>
</tbody>
</table>

* Principle: Good, Fair
Wavelength: Poor
Coherence: G

Figure 1.25 Methods of exciting the lasing medium

placed at the other focus. Without Q-switching, the laser will begin lasing as soon as the threshold is reached, and will continue until the flashlamp ceases to maintain the population inversion. The process is illustrated in Figure 1.26. For a material like ruby, the lasing in this mode is somewhat erratic and the output consists of a series of spikes that are randomly spaced in time with a mean spacing of about five microseconds. For a four level material, the output is much smoother.

A Q-switch inhibits lasing and the excitation of the material becomes more complete. The fluorescence lifetime in ruby are relatively large (milliseconds), allowing the excitation to continue as the flashlamp operates. Q-switching should be accomplished in less than the fluorescence lifetime, since further pumping does not effectively excite the material. Figure 1.26 shows that the Q-switch essentially drops the laser threshold and allows the stored energy to be swept from the cavity in a short time.
Figure 1.26 Laser output comparison with flashlamp current

A laser can be repetitively Q-switched as long as 1) the lasing material is being pumped and/or has a sufficient population inversion, 2) sufficient pumping time is given between each switching, and 3) the proper temperature and threshold conditions are satisfied. Figure 1.27 illustrates the process. The output power will depend on the level of excitation of the material. As the pulses are pushed closer than a few hundred microseconds, in typical systems, the power in the second pulse will begin to diminish, depending on how much energy had been removed from the rod during the first pulse. Commercial lasers can push these pulses to about one microsecond separation for a few pulses. Repetitive Q-switching can be done with any of the devices described above, including the chemical Q-switches, but chemical Q-switches are more erratic and harder to control than the other types.

To produce light pulses with closer spacing than one microsecond has, in practice, been accomplished by combining the pulses of more than one laser. This can also be done by dividing the cavity into more than one independent cavity as shown in Figure 1.28. Each part of the cavity can then be independently Q-switched with an arbitrarily small spacing.

High coherence and the TEM00 mode in solid state lasers is attained by aperturing the laser cavity to a beam diameter of less than 2 mm, operating the laser at near its threshold, and by using wavelength selective dyes or etalons in the cavity. All of these devices tend to reduce the total power output of the laser. To obtain both high energy and high coherence requires that the relatively weak pulse be further amplified. Figure 1.29 illustrates the most common methods for achieving such an output. One or more amplifier is added to the basic oscillator. A ten centimeter long ruby rod can amplify the intensity by about ten times. A ruby oscillator and two amplifiers can produce about two joules of highly coherent light in 20 nanoseconds.

Among the newest commercially available lasing materials of potential interest here is Alexandrite. This material can be tuned over a small range around 700 nm. Exactly how to use this to advantage is still open to the ingenuity of the readers. Another material that is showing promise is GSGG (gadolinium, scandium, gallium, garnet). This material has many of the properties of YAG and a lower threshold. The lasing wavelength is 1060 nm. The major prospect of the material lies in its potential for producing a lightweight laser, for example, for airborne applications. Finally, titanium doped sapphire is yet another new and promising material.

1.5.3.3 Gas Lasers - Much of the discussion of solid state lasers applies also to gas lasers. Gas lasers are typically of lower power output than solid lasers. They are largely pumped electronically by flowing an electric current through a plasma in a sealed tube. The conversion efficiency for converting electrical power into optical power is
Figure 1.27 Multiple Q-switching a ruby laser

Figure 1.28 Double-barrel, double-pulsing laser schematic

commonly less than one percent. The end mirrors can be fixed directly to the plasma tube for low power lasers. For higher power lasers, the mirrors are separated from the plasma tube, and the ends of the tube are sealed with windows placed at the Brewster angle with respect to the optical axis to reduce losses in the cavity. This supports lasing of polarized light favored by the Brewster windows, leading to a highly polarized output. Q-switching of a gas laser does not significantly enhance power output, but the types of light valves described above for Q-switching are used in the cavities to modulate the power output.

Typical power outputs range from milliwatts to watts. Tuned etalons are used in the cavities to enhance single frequency output when required. The most commonly used lasers are HeNe (633 nm), Argon (488 and 514 nm), and CO₂ (1060 nm). Less commonly used are HeCd (325), Kr (330 and 647), and N (337) (see Table 1.1).

1.5.3.4 Diode Lasers - Although they are not yet widely used in this field, diode lasers hold great promise for many applications because of their high efficiency and small size. Diode lasers are orders of magnitude more efficient than gas lasers. The typical output of a commercial diode laser ranges from microwatts to tens of milliwatts for a CW laser weighing only a few ounces. In the pulsed mode, such a laser can generate hundreds of watts of peak power in pulses that are 10 to a few hundred nanoseconds in duration.

The cavity of a diode laser encloses the junction region of the semiconductor between two reflective faces. The cross sectional dimension of the cavity is extremely
small and usually rectangular, commonly about one by ten micrometers. This leads to a highly divergent beam because of diffraction, the divergence being its greatest in the short dimension direction. So the output beam must usually be conditioned with lenses to make it usable.

In applications where a small light source is required, a diode laser is often an ideal choice. One current limitation with diode lasers is their long wavelengths, starting at about 780 nm. This requires special detectors and prevents their use with photographic emulsions. Recently, diode lasers in the green have been developed, but they are not yet commercially available. This development could be extremely important for laser diagnostics. In other new developments, phased diode arrays are being produced with net output power capabilities of watts CW.

1.5.3.5 Metal Vapor Lasers - In recent years, copper and gold vapor lasers have become commercially available and they fill an important gap in flow diagnostics. They emit light in nominal 20 nanosecond pulses at pulse rates up to a few thousand per second at 510 nm for copper and 628 nm for gold. This makes them excellent candidates for use in high speed photography, schlieren and shadowgraph. Unfortunately, they have extremely poor coherence.

1.5.3.6 Excimer Lasers - The word “excimer” is extracted from the term excited dimer. A dimer is an unstable molecule formed at high temperature from two or more molecules such as a noble gas and a halide (for example, xenon and fluorine). Flashlamp pumped excimer lasers operate in the ultraviolet at quite high pulsed output powers. With pulsewidths of tens of nanoseconds and energies comparable to ruby and YAG, they fill an important gap in flow diagnostic applications. They are commonly used as a pumping source for dye lasers.

1.5.3.7 Dye Lasers - Dyes are extremely complex molecules, characterized by many possible energy transitions. Consequently, when a dye solution is used as a lasing medium, many different wavelengths are possible. With a suitable combination of dyes, it is possible to produce a laser that is tunable almost continuously across the entire spectrum. A dye laser comprises a dye cell through which the dye solution flows, a pumping source which may be either a flashlamp or another laser, a wavelength selector which may be a prism or grating, and frequency selecting etalons to further enhance a narrow band frequency output. When another laser is used to pump a dye laser, the wavelength of the dye laser can be no shorter than that of the pumping laser. Therefore, excimer lasers are often used to pump dye lasers, since the resulting “single pass” dye laser can produce light over a wavelength range from ultraviolet to infrared.

1.5.3.8 TEA Lasers - TEA is an acronym formed from transversally excited atmospheric. Commonly operated at atmosphere with CO₂ as a lasing medium and in the pulsed mode, these lasers can fill an important gap where high power, short pulses of light at 1060 nm wavelength are required. Pulselwidths range from nanoseconds to microseconds, and powers range as high as hundreds of kilowatts.
CHAPTER 2
USING LASERS WITH CONVENTIONAL OPTICAL TECHNIQUES

2.0 BACKGROUND

Lasers are often better light sources than more conventional ones, but they do suffer some disadvantages that must be traded off against advantages. An arc source, flashlamp or even an incandescent source can often outperform a laser depending on what is required. Therefore, it is important to choose a light source because of its features and mainly because it is a laser.

In this chapter we describe applications of lasers in photography, shadowgraphy, schlieren, interferometry, and deflectometry, emphasizing those applications where lasers extend the methods beyond their usual range of applicability. These will illustrate how the unique properties of laser light can improve space and time resolution and improve the ability to separate signal from noise. Coherence, intensity, and controllability associated with lasers can be used individually or in combination to achieve such improvement.

Coherence - Temporal coherence implies single wavelength. An event can be illuminated with laser light and observed or recorded through an extremely narrow bandwidth optical filter that highly attenuates all but the laser wavelength. Therefore, experimentation can be extended into cases involving much greater ambient light intensity. Additionally, high coherence simplifies both the theoretical description and the practical use of interferometry and has resulted in a widespread increase in its applicability. On the other hand, spatial coherence makes possible the production of extremely small, high-intensity concentrations of light not previously available in optics, significantly improving the spatial resolution limits of optical probes.

Intensity - Extremely high-intensity illumination can be produced from a laser. This is, of course, beneficial when high-intensity is needed, but this should not be confused with total energy. With few exceptions, lasers are not efficient energy sources. Flashlamps, arc sources, induction sources, and other forms of energy delivery should be considered equally as candidates where the task is the delivery of energy or the production of large quantities of light.

Controllability - Spatial coherence, temporal coherence, and polarization, all make laser light much easier to control in space and time than incoherent light. For example, almost all precision optical elements are easier to design for a single wavelength including reflectors, lenses, shutters, filters and modulators. Also, highly collimated light producible with spatially coherent light can be transported over large distances without losses. Further enhancing controllability are electro-optical modulators and shutters that can create pulses of light with great precision in time, placement and duration.

Disadvantages - The problems of speckle, interference, and diffraction noise often render recordings made with highly coherent light extremely difficult to use. Unwanted diffraction patterns, "parasitic" interference fringes, and patches of light can mix coherently with data making the result an even more serious degradation than simple addition of intensities. In general, therefore, laser optics must be chosen more carefully. Scratches and flaws are more deleterious, and cleanliness is much more important with coherent light. It has been said that if natural light had been coherent, then the discovery of incoherent light would have represented a major breakthrough.

2.1 Laser Photography and Videography

Photography and videography in aerodynamics can be divided into the following tasks:

1. Illuminating the object field,
   a. with light of a desired structure,
   b. with the required intensity,
   c. from the desired direction,
   d. for the required time,
   e. at the desired time.

2. Recording the light that has passed through or reflected from the object field,
   a. at the required time, possibly synchronized with other events,
   b. with the required temporal resolution,
   c. with the required spatial resolution,
   d. while rejecting other light sources that are of no interest.

Section 1.4 described the various methods available to control coherent light, and Section 1.5.3 described the characteristics of available lasers for producing the light. This section will show by example how the methods can be combined to achieve the above with some rather spectacular results. The most advantageous uses occur when one of the above nine tasks is at an extreme. Higher intensities, more complicated structures, delivery at larger distances, with greater ambient light rejection, and more precisely timed delivery (cham can be derived from nonlaser sources) are extremely useful features of lasers when photographic limits are approached.
2.1.1 Photography of Hypervelocity Objects

Studies in aeroballistic ranges sometimes require accurate in-flight measurements of the position, contours, and surface conditions of high velocity models. Figure 2.1 illustrates a system in use at the AEDC Range G1,2 exemplary of the techniques of laser high-speed photography. In this case, the laser light is first diffused by a ground glass to reduce the effects of diffraction. Then it is concentrated by a lens into a region of interest where it finally strikes a background screen. Six cameras, each having four lenses that are focused at a different depth in the field, cover the sample volume. Narrow-band filters positioned in the image plane remove the ambient light.

![Diagram of Front-light/back-light laser photographic system](image)

The specifications of the components in the system vary according to test conditions. For example, narrow-band filters range from one to 1000 angstrom bandwidth depending on model luminosity. Red sensitive films include two major types, Kodak S0243, having a resolution of 500 line pairs per mm and an ASA rating of 1.6 and Kodak Linagraph Shelleburst, having a resolution of 100 line pairs per mm and an ASA rating of 125. In the system described here, the higher speed film must be used when F/numbers greater than 45 are required.

The lighting procedure used here is intended primarily to produce a front lighted photograph of the model in flight. This requires an advance signal to fire the ruby laser at the proper time. The first signal, provided to the laser when the model is still far upstream, fires the laser flashlamp at about one usec before the model arrives. A CW laser beam interruption then identifies the arrival of the model near the sample volume whereupon a firing signal Q-switches the ruby laser. Even if the model misses the front-lighted volume, it may still traverse a region where it is backlit. Clearly, the problem would be made easier if the exact position of the model were known apriori. Figure 2.2 includes a series of laser photographs to illustrate the types of data that can be produced in such a system as described here. Figure 2.2a shows the appearance of a model in flight at various downrange stations showing the change in surface condition and contour.

Example: The 5000 meter/sec model shown will move 100 micrometers during the 20 nanosecond exposure, representing a blurring of the resulting photograph.

Figure 2.2a shows the laser photograph of a model in which a coolant is ejected from the nosetip in flight. Figure 2.2c illustrates the photograph of a model in which the background is striped to augment the visualization of flow.

Other variations of the laser station described above provide for specific diagnostic needs. For example, four lasers were combined to provide high-speed photographic sequences such as shown in Figure 2.3. These are valuable when a rapid time development takes place. Another example shown in Figure 2.4 is the combined front-lit photograph of the model surface and the schlieren photograph of its flow field.
Figure 2.2a. Erosion model in the 300 meter range. Top figure is early in flight. Bottom figure is down range. Note eroding nosetip, and impact craters.
Figure 2.2b. Transpiration cooled nosetip in the AEDC track G. Note coolant material flowing from the nosetip of the model, which is traveling at 5000 meters/second. The striped background aids in viewing the flowfield.

Figure 2.2c. Front-light photography of a hypervelocity projectile.

Figure 2.2 Front-light laser photograph of hypervelocity models produced in AEDC hypervelocity ranges and tracks.

2.1.2 Videography of High-Speed Events

The standard television scanning rate is 30 new pictures per second, a speed that has been chosen to satisfy viewing requirements. Such a slow speed has always severely restricted the use of TV in high-speed recording. One way to partially circumvent the slow-scanning race is to use a camera that retains a latent image momentarily, produce the image with a short burst of light from a strobe or a laser and, thus, stop the high-speed motion at a still limiting framing rate of 30 per second. New developments in videography in the past decade, however, have extended the ability to frame at higher rates and to control the entire image recording process in many useful ways. Consequently, videography has begun to play an ever-increasing role in optical diagnostics.
Figure 2.3

Four laser camera capturing the impact of a hypervelocity model with spherical particles. The model surface is five centimeters across and the spheres are three millimeters in diameter. Photos were made in AEDC hypervelocity track G.
Figure 2.4 Combined front lit and laser schlieren photograph of a 5000 m/sec model in 250 torr air. Note the ability to see boundary layer transition on the model.
The most significant factor is the development and commercial availability of charge coupled devices (CCDs). These sensing elements are configured in linear and two-dimensional arrays. The typical array comprises 15 micrometer square sensors spaced by 45 micrometers center-to-center. Linear arrays containing up to 2048 detectors and two-dimensional arrays up to about 300 by 400 are available commercially. The arrays are exposed in one step, leaving the intensity information locked within each detector. Then the array is scanned electronically in a serial fashion producing a string of numbers or a set of strings of numbers that describes the intensity distribution. This process provides a very natural way to sample and digitize the information in an intensity distribution, casting it into a form that is optimum for controlling and processing by computer.

The remarkable development of these devices owes much to the simultaneous development of small, inexpensive, but powerful computers. What is a fairly simple control process for a small computer today would have been prohibitive in cost and complexity for most laboratories only a few years ago. In this sense it is hardly meaningful to separate the sensor and the computer, since the computer ultimately takes the string of numbers, performs operations on it and then prepares it for presentation on a monitor or printer, possibly as an image or in some even more useful form.

An exemplary application is in the study of the effect of aerodynamic flow fields on imagery (aero-optics). This might be encountered, for example, in astronomy or in photography from an aircraft. Figure 2.5 describes an experimental setup. A diffraction limited point image is focused on a two-dimensional array CCD after having traversed a flow field. The camera system in this case was manufactured by the Spin Physics Division of Eastman Kodak Company. The entire array can be scanned at rates up to 2,000 frames per second. The data is stored on video tape and can be played back at adjustable rates for viewing. This provides the capability to examine instant replays of relatively high-speed events.

Figure 2.5 Conceptual experiment for measuring aero-optic effects. The incoming wave front is distorted by the flow field, limiting the attainable resolution of an imaging system in the vehicle.

Figure 2.6 illustrates the resulting data showing how the point source changes its shape under the influence of an aerodynamic flow field. As can be seen here, the image has been processed by the computer to find the size and centroid of the image automatically. The image increases in size (blur) as a result of the phase modulating effect of small scale turbulence and the centroid moves about (jitter) as a result of the larger scale turbulence. Models for the aberration of such images have been developed and have been partially confirmed in these types of studies. CCDs offer extremely good versatility in image evaluation. For example, higher scan rates can be attained by scanning a smaller part of the array, trading off data per picture for a higher framing rate. The Spin Physics camera can frame at rates up to 12,000 frames per second. Furthermore, at lower rates it is even possible to process and display processed pictures in near real time.
Figure 2.6 Aero-optical manifestation in an image. Vibration moves the image around; large scale turbulence and aberrations distort the image; small scale turbulence tends to diffuse and blur it more severely. The time over which an image is observed affects the image differently for each case.

Still higher recording rates and faster shuttering times can be provided by taking advantage of short laser pulse lengths and scanning optics. A CCD array can be divided into sectors that are exposed sequentially with a pulsed laser that is electro-optically scanned between the sectors. Various examples of the lasers described in Chapter One could provide exposures of less than 20 nanoseconds and repetition rates up to one megahertz.

The field of CCD/computer imagery is still young and offers much promise for the diagnostician in aerodynamics. Also, this is likely to be the source of information handling capability that will be needed for the ever increasing amount of data generated by optical diagnostic systems.

2.2 Structured Light Methods

Of the three classes of lighting that have been discussed, direct, diffuse, and structured, the most elegant is structured light. When structured light is used, the sorting or processing of information actually begins when the field is illuminated. This is accomplished by virtue of the function that is structured into the light. The following are some of the structures that have been employed and will be discussed in the following sections.

1. Thin sheet-of-light, also called laser planing and screening.
2. Two or more lines of light.
3. Two or more sheets of light.
4. Sinusoidal intensity distributions of light.

2.2.1 Sheet-of-Light Flow Visualization

The visualization of complex three-dimensional flows is especially difficult with ordinary means, because the planes that are deep within the flow field are masked by the ones that are closer to the viewer. This difficulty is partially overcome with light sheet flow visualization by illuminating only the plane in which the flow is to be viewed. The typical laser sheet-of-light is generated directly from the laser beam by passing it through a cylindrical lens. The resulting sheet is typically one mm in thickness and diverges in thickness at the same rate as the original laser beam (commonly about one milliradian) and diverges in width at a rate determined by the focal length of the cylindrical lens. Lasers have replaced arc sources in most facilities using this type of diagnostic tool because it is possible to generate thinner sheets of higher intensity light.

Figure 2.7, showing a light sheet system employed in an ONERA wind tunnel, is a typical system geometry. The sheet generating optics is usually mounted on a traversing system that allows the recording of adjacent planes in the flow. The flow itself must be seeded with light scattering material such as smoke. Light scattered from the plane can be recorded in the form of movies or stills depending on the phenomena under study. The laser is usually a CW laser having a power of 100 or more milliwatts. Figure 2.8 is a sample set of data from the ONERA system showing the flow field in the wake of an
Figure 2.7 Laser sheet flow visualization configuration of ONERA aircraft model. This method is extremely graphic in its ability to show vortex structure, because the scattering particles are actually spun out of the vortex leaving a dark region where the vortex is centered. Other flow features also shown by sheet-of-light viewing include the transition from laminar to turbulent flow, shock waves, boundary layers, and separating flow.

2.2.2 Particle Image Displacement Velocimetry (PID)

Velocimetry, accomplished with the use of tracers in the flow, is one of the oldest velocity measuring methods. The technique consists simply of tracking the position of a tracer at two or more times and then drawing the associated velocity vector. The use of a sheet of light makes this a much more tractable method in almost any flow. The scatter center (SC) must remain within the sheet for the measurement. The sheet is pulsed twice or more to record the time separated positions of the SCs or alternatively a camera can be shuttered in some other way to produce the two recordings. Surprisingly, the newest refinements of this method represent a state of the art in velocity measuring techniques. The new method, sometimes called speckle velocimetry, is actually a multiply exposed image that can be free of speckle, and particle image displacement velocimetry is more correct terminology.

Figure 2.9a from the work of Meynart shows a multiple exposure image of a laser sheet produced in a convective flow with an illuminating apparatus similar to that of Figure 2.7. The SCs in this flow are 5 micrometer particles. Ten exposures of 20 msec duration separated by 3 second intervals were made. If any part of this photograph is illuminated by a small laser beam then a set of Young's fringes will appear in the space beyond the photograph. The fringes will be oriented normal to the flow direction and will be spaced inversely proportional to the speed in the flow. It is possible to determine the local velocity component in the plane in this manner. In fact, the same methods described in Chapter 3 for analyzing interference fringes can be used to automate the process. In Figure 2.9b-c Meynart has also shown an alternate technique for processing the data. By Fourier transforming the image and spatial filtering, the isospeed contours result.
(a) The light sheet is parallel to the model surface showing a lengthwise cross section of the vortex created by the edge of the model. Note the vortex beginning to collapse at the extreme right.

(b) The sheet of light is perpendicular to the model surface. Four recordings are superimposed showing cross section of the vortices at various stages of development.

Figure 2.8 Sheet of light flow visualization of a vortex shedding from a delta wing.

This technique has been restricted so far to relatively low velocity flows with fairly large SCs. It has been used in gas and liquid flows up to a few meters per second. Figure 2.10 from the work of Lourenco et al. was produced with a 7-watt argon laser operating at the 514 nm wavelength and shuttered with a Bragg cell. The tracer particles were titanium dioxide, having a diameter range between 10 and 30 micrometers. The film used here was Kodak Royal-x having an ASA rating of 1250. Figure 2.10a shows the overall view of a seeded vortex. Figure 2.10b is the result of processing a part of the image to produce the velocity vectors. This was produced by passing a small probe beam through the negative, generating the Young’s fringes, processing the fringes using techniques described in Chapter 3, and superimposing the resulting vectors on the original picture.
(a) Multiple exposure photograph showing cross section of the toroidal flow.

(b) Filtered image with fringes displayed with reverse contrast, vertical velocity component. Contour differences are 15.5 micrometers/sec.

(c) Filtered image horizontal component, contour differences are 18.8 micrometers/sec.

Figure 2.9 Laser sheet velocimetry of a convection torus in glycerin.

A variety of techniques have been reported to improve data quality. For example, even though the original figure is produced on a fast, large-grained material, improvement in the quality of the fringes results if a secondary contact print is produced on a finer grained material. This tends to reduce the film grain noise and is an illustration of a case where the film grain causes a more serious problem with coherent light than with incoherent light. Young's fringe contrast also is improved with the number of exposures up to about 10 exposures.
Figure 2.10a. Light Sheet Velocimetry - Observe the four exposures of scatter centers in the vortical flow field.

This technology is quite active and appears to have tremendous potential for new application and refinement. For example, the method may be applicable to a large number of gaseous flow problems if the proper pulsed laser is applied. The automation of data handling should also make this a much more useful method.

2.2.3 Deflectometry and Hartmann Methods

As described in Chapter One, when light passes through a flow field containing refractive index gradients, a refraction or bending of the light rays occurs in the direction of the gradient. This can equivalently be described as a phase shifting of the wavefront; the wavefronts are being compressed more in the region of higher refractive index. To determine the effect of the flow field on the wavefront, one can measure either the amount of deflection of the rays or the amount of phase shift. The former is called deflectometry while the latter is called interferometry.

Deflectometry produces a direct measure of the actual deflection of individual bundles or rays of light passing through a flow field. The light structuring to produce the individually separable rays can be done in one of several ways. Figure 2.11 illustrates one such way in which an aperture (Hartmann) plate is placed immediately following the object field. The plate comprises an array of holes ranging in diameter from about one mm to one cm. In some cases the apertures are actually lenses in the
plate that focus the rays of light onto position sensing detectors on the receiving side. These constantly monitor the centroid of the light bundle. The emergence of solid state CCD arrays and position sensors have made this technique tractable in many flow diagnostic applications.

The sensitivity of a Hartmann system depends on how accurately the sensor can locate the ray centroid and on the distance from the Hartmann plate to the sensor. Typically such a system is capable of resolving about one microradian of deflection of the light ray. This is quite remarkable sensitivity. It corresponds to about one-fiftieth wavelength increase in pathlength per cm across the flow field.

If a wavefront is known to have structure that is larger than the aperture spacing, the Hartmann system can be used to define the wavefront. However, caution must be used in interpreting results when the spatial frequency content of the wavefront is not so bounded. This is a sampling system and is governed by the nuances of sampling theory. If the wavefront is undersampled, the resulting data can be plagued with misleading results.

Figure 2.11b describes moiré deflectometry. In this case the light waves are structured by grids, strategically placed at one or more locations forming parallel sheets of light. If a grid is placed on either side of the flow field, then without flow a linear set of moiré fringes will result, depending on grating frequencies and
Figure 2.11 Hartmann and moiré methods of measuring the angle of refraction.

orientation relative to each other. A flow field has the effect of optically distorting the grid preceding the flow field. When the distorted grid is observed against the undistorted grid, the moiré pattern is appropriately distorted. This arrangement does not require the light to be collimated; it requires only that both grids remain in focus in the viewing system.

A different type of deflectometry results if the two grids are placed after the flow field. Here, collimated light must be used. The moiré pattern forms because the light, having been deflected, passes through the two different grids at different places and will be blocked by the second grid differently than light that has not been deflected.

In moiré deflectometry laser light is usually not advantageous because temporal coherence creates significant diffraction noise. It would appear that some of the lasers having poor temporal coherence but good spatial coherence such as copper and gold vapor lasers might prove useful in this type of application.

These methods have not seen widespread use as of yet so it is not clear how important they will become in flow diagnostics. They offer the advantage of adjustable sensitivity (by rotating the grids), an extremely broad dynamic range, and utility in very strong flow fields. Perhaps with the combination of new videography methods, new lasers, and new analytical techniques, they could become powerful diagnostic tools.

2.3 Shadowgraphy and Schlieren Methods

Shadowgraphy and schlieren methods are the oldest forms of flow visualization of phase objects and, consequently, the methods are highly refined. It is extremely difficult to produce with laser light the same cosmetic quality that is routinely produced with conventional light sources. Even so, because shadowgraphy and schlieren methods can be performed so easily with lasers and in conjunction with holography and interferometry, they are used extensively in flow diagnostics.

Figure 2.12 illustrates the two methods. In shadowgraphy, the light striking a refractive index gradient, is refracted leaving a darkened zone. The recording must be made in an out-of-focus plane to achieve significant sensitivity. This is one of the disadvantages of shadowgraphy. When laser light is used, diffraction effects can be especially annoying. Schlieren methods remove the problem of out-of-focus images by enhancing the image of the refractive index gradient in a totally different way. The collimated wave that has passed through the field of interest is focused by a lens or mirror onto a filter plane. For the simple knife-edge filter case, the process can be modeled simply as the blocking of the refracted ray in the filter plane. Consequently, the part of the image containing refractive index gradients appears dark in the final image if the direction of the gradient is normal to and in the direction of the knife edge.

There is a significant difference in the operation of the filter plane depending on whether coherent or incoherent light is used. When coherent light is used, this plane represents the two-dimensional Fourier transform of the amplitude distribution in the field of interest. This makes for a wide range of possibilities for performing mathematical operations on the input image before it is finally viewed. Some of these were analyzed in detail in AGARDograph 166. Since they have ostensibly not reached any significant use in flow diagnostics (to the author's surprise), they will not be repeated here. Such operations include differentiation of the image, Hilbert transformation, and phase contrast interferometry. Regardless of this potential capability, both shadowgraph and schlieren methods are almost exclusively used for qualitative work.
2.3.1 Space Shuttle Schlieren System

The NASA Spacelab Fluids Experiment System (FES) for the study of fluid dynamics and crystal growth in zero gravity is equipped with both holographic recording and a laser schlieren system. A 20 Mw HeNe CW laser is used and the resulting images are monitored and recorded on standard TV. The knife edge is adjustable along the optical axis and can be rotated about the axis. This system has been used in orbital experiments and will be used as an example here for typical schlieren system interpretation.

In Spacelab 3 experiments, a triglycine sulfate crystal was grown in zero gravity to determine the quality and feasibility of growing crystals in a totally diffusion controlled process. (In a gravitational field, the process is heavily influenced by natural and forced convection.) By inserting a seed crystal into a saturated solution, the process could be regulated by controlling the temperature difference between the fluid and the crystal. When the crystal temperature is significantly higher than that of the fluid, the crystal dissolves. As the temperature of the crystal is reduced, a critical temperature is reached, depending on solution concentration, where the crystal begins growing.

In this case, refractive index increases with concentration of solute. When the crystal is growing, the solution immediately next to the crystal becomes depleted and the refractive index gradient is positive in the direction normal to and away from the crystal face. When the crystal is dissolving, the reverse is true. The schlieren system served two major functions:

1. To determine when the crystal was growing and when it was dissolving and,
2. To determine if there were any residual convection currents in the cell.

Figure 2.13 represents the system geometry and the matrix of possible experimental conditions. The diagnostics can be optimized by adjusting the knife edge according to the experimental condition. The knife edge should be positioned with its edge normal to the refractive index gradient so that the refracting ray (which bends in the direction of the refractive index gradient) strikes the knife edge. This requires rotating the knife edge by 180 degrees when the process changes from dissolution to growing. The growth of other faces of the crystal, furthermore, requires a rotation of the knife edge to maximize sensitivity for that particular refractive index gradient.

2.4 Interferometry

The basic relationships for most interferometers were derived in Chapter One. In this section the most common types of interferometry are reviewed and some of the applications in aerodynamics are described. In the last section, the effects of a flow field on a transmitted wavefront were determined by measuring the amount of the deflection of the rays of light that define the wavefront. Interferometry determines this effect by measuring the variations in optical pathlength through the flow field or equivalently by determining the amount of the phase shift of the wavefront impressed upon the wavefront by the flow field. In interferometry it is assumed that the bending of the rays in the wavefront are negligible, that they pass straight through the field with a phase shift only.
Figure 2.13 Appearance of schlieren image for various knife edge settings and density gradients.

Where refractive index gradients are so great that ray bending is not negligible, the data interpretation near such a region is extremely complex. Some attempts at analyzing such data have been made but no general solution of the problem has emerged.

Interferometers find uses in aerodynamics in many more applications than flow diagnostics. An extensive treatment of such applications is outside the scope of this work; however, some of the basic configurations are mentioned briefly.

2.4.1 Michelson Interferometer

The Michelson interferometer is a widely used instrument which in its simplest form employs a single beam splitter and two mirrors (Figure 2.14). Variations are used in phase and distance measurement, measurement of refractive index, and measurement of optical frequency change.

Vibration Measurement

Commonly used to examine vibrational stability of stable tables this apparatus can be assembled in minutes even with poor optics. If the $k$ vectors are accurately aligned, Equation 1.10 is applicable. Motions of either $M_1$ or $M_2$ causing a $\Delta L = \Delta(L_2 - L_1)$ results in a change of phase difference $\Delta \phi$ at the origin, $O$, which is given by

$$\Delta \phi = \frac{2\pi \Delta L}{\lambda}$$

Thus, a change in $\Delta L$ by $\lambda/2$ causes the observed output to pass through an entire cycle of intensity. Likewise, a change in the refractive index, $n$, in either path causes a shift in $\Delta \phi$ allowing the detection of air currents.

Rotation of either mirror misaligns the $k$ vectors and Equations 1.8 or 1.10 are applicable. For example, as previously stated, a 10 second rotation of $M_1$ produces a set of linear fringes spaced by one cm ($\lambda = .5$). Linear fringes can be obtained only if all optical elements are flat.

Testing Optical Surfaces

If the optical components are flat, one can test mirrors, lenses, and windows. A window is tested for surface flatness by inserting the window into either leg of the interferometer. Fringes will be initially straight. If both surfaces are flat, they will remain straight. If the surfaces are parallel, the fringe frequency will remain unchanged with or without the window present. The method is useful in testing flatness to about $1/10$ and parallelism to about one second. The individual surfaces can be tested by replacing $M_1$ or $M_2$ with the surface to be tested.
Figure 2.14 Michelson interferometer.

**Measurement of the Acceleration of Gravity**

A corner cube reflector reflects light antiparallel to itself. The device shown in Figure 2.15 has been used to accurately observe the velocity of such a prism in free fall. Equation 1.18 applies. Assuming $u_0 = u_0' = u_0''$

$$I(x,z,t) = 2u_0^2 \left[ 1 + \cos \left( \frac{4\pi}{L_1} (L_2(0) - L_1 - gt^2/2) \right) \right]$$

(2.2)

Example: The position of the prism can thus be determined at any time with an accuracy of better than, say 1/20. At some time $t'$, the signal at the photo sensor is oscillating with a frequency, $f' = gt'/L$. At $t' = 1$ sec, $g = 1000$ cm/sec$^2$, $x = .5$ microns then $f' = 20$ megahertz. It should be mentioned that the instrument can be considered a Doppler measuring device, wherein the light beam in the $L_2$ leg is Doppler shifted (in the example given, by 20 MHz). As such, Equation 1.7 applies giving an equivalent result.

Figure 2.15 Measurement of the acceleration of gravity.

**Measurement of Coherence Length**

Formation of fringes at the observation plane requires coherence of the input beam. Fringe contrast is defined for ideal interference as

$$C = 2u_0' u_0''/(u_0'^2 + u_0''^2)$$

(2.3)

Associated with every light wave is a coherence length. If the quantity $\DeltaL$ is increased from zero, one will find that the fringe contrast is reduced below that expressed in Equation 2.3. Coherence length is commonly defined as $(\DeltaL)^2/2$ where $(\DeltaL)^2$ is that difference which reduces $C$ to zero. Coherence lengths vary from a few cm for ordinary lasers to many meters for specially designed lasers.

2.4.2 Parallel Plane Interferometers

A large variety of interferometers employ two or more nearly parallel reflecting planes to add beams that have reflected from the individual planes. Some of these are extremely simple and yet useful. Many of these were not practical devices before the advent of the laser because they employ unequal pathlengths or spatial mismatching of the split beams.
A simple measurement of the coherent length of a laser can be accomplished with one beam splitter mounted parallel to the front reflector of the laser (Figure 2.16). The front laser reflector provides the second leg of the interferometer. When the distance D is one-half the laser coherence length the fringe contrast will be very poor. The interferometer can also be used for observation of small variations in the product nD.

![Figure 2.16 Parallel plate interferometer.](image)

**Plane Parallel Plate**

A plane parallel plate (etalon) is commonly employed in a laser cavity to spoil the cavity for all frequencies except a discrete set. The net reflectivity of intensity of such a plate having reflectivity R at each face is given by (assuming a perfect plate, no losses) (Figure 2.17)

\[
R_p = \frac{4R \sin^2 \left( \frac{2\pi nh \cos \phi}{\lambda} \right)}{1 + 4R \sin^2 \left( \frac{2\pi nh \cos \phi}{\lambda} \right) (1-R)^2}
\]  

(2.4)

A plane wave will not lose any energy to reflection if \( m = 2\pi h / \lambda \), where \( m \) is an integer.

**Example:** \( \phi = 0, h = 0.5 \) cm, \( \lambda = 0.5 \) \( \mu \)m. Transmission peaks occur at approximately every 0.25 angstroms, for all wavelengths satisfying \( \lambda = 2\pi h/m \).

Such a plate can be inserted into a laser cavity and tilted until its resonant wavelength matches one of the longitudinal modes of the laser. The only losses at that wavelength are then caused by scattering (not reflection) from the plate, and by the drift of the laser wavelength away from resonance. The remaining longitudinal modes may not all be eliminated entirely but they are reduced considerably in intensity below the resonant frequency. The coherence length of a laser can be increased in this manner with a very inexpensive piece of glass; however, commercially available etalons for this purpose are generally of very high quality and are expensive.

The transmissivity of such a plate is given by

\[
T_p = \frac{1}{1 + 4R \sin^2 \left( \frac{2\pi nh \cos \phi}{\lambda} \right)}
\]  

(2.5)

A typical plot of the transmissivity versus \( 2\pi h \cos \phi / \lambda \) is sketched in Figure 2.18. As can be seen, as the reflectivity of the surfaces approaches one, the plate becomes more and more selective in transmitting specific wavelengths.
Fabry-Perot Interferometer

Figure 2.18 Transmissivity of a parallel plate.

With the discussion above, the remaining is simply a discussion of hardware and terminology. A Fabry-Perot interferometer consists typically of two plane parallel reflecting surfaces with an air spacing \((n=1, \text{ Equation } 2.5)\) and a variable separation, \(h\). The variation in \(h\) is usually accomplished by (1) thermal expansion of the spacers, (2) pressure control in the air space, (3) spacers constructed of piezo-electric crystals that can be expanded electrically. The familiar terminology for such an interferometer includes:

(1) **Free Spectral Range** - The interferometer has a transmission peak in every free spectral range of frequency. It has a peak transmissivity for many different frequencies. From Figure 2.18 we see that this is given by:

\[
\Delta \nu = \frac{c(\text{velocity of light})}{h \cos \theta'}
\]

(2.6)

This is typically a few gigahertz or a wavelength spread of a few hundredths of an angstrom.

(2) **Finesse** - Usually defined as the free spectral range divided by the width of the transmissivity curve at half maximum, or

\[
\text{Finesse} = \frac{\Delta \nu}{\text{transmissivity width at } 1/2 \text{ max}}
\]

(2.7)

Finesse varies from zero to infinity and is a measure of the instrument resolution. The free spectral range is essentially divided into \(F\) resolution elements. The maximum achievable finesse in practice is about 200.

(3) **Peak Transmission** - Practical instruments are not loss free. Peak transmission is defined as maximum transmissivity, \(T_p\), for a real interferometer.

(4) **Contrast Factor** - Peak Transmissivity/Minimum Transmissivity

To this point we have discussed only parallel illumination. If a diverging (point source) illumination is employed, one observes a set of circular fringes for each wavelength. The transmissivity repeats itself each time \(nH \cos \theta'\) increases by one-half wavelength.

Other variations of the interferometer employ spherical reflectors or one spherical and one flat reflector. Laser cavity mirrors in this respect constitute a Fabry-Perot interferometer.

2.4.3 Mach-Zehnder Interferometer

Figure 2.19 represents the geometry of the widely used Mach-Zehnder interferometer. Two laser beams are generated by beam splitting. One beam is directed over a path which is uniform in refractive index (Path 1) and is subsequently used as a reference for the interferometer. The second beam passes through the interest field before it is mixed with the reference. The interest field is imaged to an observation plane. The mixing of the two waves characterizes the refractive index changes through changes in the interference fringes.

**Infinite Fringe Interferometry** - With no disturbance in the interest field the linear fringes observed at the output can be adjusted out by precisely aligning the \(k\) vectors (assuming high quality optics, windows, mirrors, and beam splitters). This results in essentially a single fringe which is infinitely wide. The common practice in explaining a resulting fringe pattern when a disturbance is present is to apply an approximation of Equation 1.7 which reduces it to a form similar to Equation 1.18. Assume that beam two passes through the interest field. Refractive index gradients change its direction resulting in a nonzero value of \(\phi_2\). Equation 1.7 predicts the proper set of fringes only if the wave remains planar, since the equation is valid only for plane wave interference.
Wave number two is, however, in general, modulated by the interest field into an extremely complicated wave. Wherever the angle \( \theta_2 \) is extremely small, it can be neglected if one still accounts for the phase shift. We can apply Equation 1.7 if the term \( kx \sin \theta_2 \) is replaced by \( \Delta \phi(x) \), since, in general, \( \theta_2 \) will be a function of \( x \).

\[
I(x) = u_{01}^2 + u_{02}^2 + 2u_{01}u_{02} \cos [\Delta \phi(x) + \Delta \phi_0]
\]  

(2.8)

The term \( \Delta \phi(x) \) represents the increased amount of phase shift suffered by the ray passing through the disturbed field. When the interest field is undisturbed \( \Delta \phi(x) \) is zero and \( \Delta \phi_0 \) is the undisturbed phase difference of the two mixed waves. \( \Delta \phi(x) \) can be calculated as

\[
\Delta \phi(x) = \int_{z_0}^{z_1} (n(x,z') - n_0)dz'
\]

(2.9)

where \( z_0 \) and \( z_1 \) bound the interest field.

The integral in Equation 2.9 defines the optical pathlength difference between disturbed and undisturbed interest fields. A particular fringe (constant intensity lines), therefore, locates contours of equal pathlength through the field. The simplest case occurs when \( n \) is not a function of \( z \). Equations 2.8 and 2.9 reduce to

\[
I(x) = u_{01}^2 + u_{02}^2 + u_{01}u_{02} \cos \left[ \frac{2\pi}{\lambda} [n(x)-n_0] (z_1-z_0) + \Delta \phi_0 \right]
\]

(2.10)

In advancing from one fringe to the next, the optical path length increases (or decreases) by \( (n(x) - n_0) \lambda' = \lambda \). Since \( n_0 \) and \( \lambda' \) are known constants, calculation of \( n(x) \) and application of Equation 1.36 leads to density \( \phi(x) \). The function \( G \) depends mildly on temperature and upon the gas mixture. For gases, it is not significantly dependent upon wavelength.

For air \( G = 1.07 \times 10^{-22} \text{cm}^3 \). For \( \Delta z = 20 \text{ cm}, \lambda = 5 \text{ microns} \), a one fringe increase represents a density change of \( 2 \times 10^{17} \) molecules/cm\(^3\). This gives the measurement resolution in the order \( 10^{-3} \) to \( 10^{-2} \) atmospheres at room temperature.

The next simplest case is that of the axisymmetric body, namely \( n(x,z) = n(x^2+z^2) \). Equation 2.9 reduces to (Figure 2.20)
\[ \Delta \Phi(r) = \frac{2 \pi}{k} \int_{-\infty}^{\infty} \frac{n(r') - n_0}{r^2 - r'^2} r' \, dr' \quad (2.11) \]

Solution of this integral is known as radial inversion. The procedure amounts to division of the circle into circular shells in which \( n \) is assumed a constant. The integral is then solved first for the outer shell, then for the first two outer shells then three and so on, (like peeling an onion). A variety of computer programs and data reduction procedures are available.

**Figure 2.21** is an interferogram of this type. The high pressure (high refractive index) region behind the bow shock of a projectile in flight manifests itself in a fringe system that can be analyzed through application of Equation 2.11.

**Finite Fringe Interferometry** - The method just described is not always useful when the net phase shifts in the interest field do not exceed \( \pi \). It is useful to misalign the original \( k \) vectors to provide a set of reference fringes. Again, we make an approximation that the change in \( \alpha_2 \) is small and apply Equation 2.7 giving

\[ I(x,z) = u_0^2 + u_{02}^2 + 2u_{01}u_{02} \cos k [z (1 - \cos \alpha_2) - x \sin \alpha_2 + \Delta \Phi(x) - \Delta \Phi_0] \quad (2.12) \]

A linear system of fringes exists on the interferogram which, according to Equation 2.12 are equally spaced by \( 1/\sin \alpha_2 \) when no disturbance \( \Delta \Phi(x) \) is present. The presence of a disturbance changes this spacing by a fraction of a fringe, \( p \)

\[ p(x) = \Delta \Phi(x)/2\pi \quad (2.13) \]

**Figure 2.21** Three fringe interferogram of a Cal. 30 cone cylinder at Mach 2.5 in air at one atmosphere pressure.
The quantity $p(x)$ is measured and Equation 2.12 is applied to Equation 2.9 (or 2.11 for axisymmetric fields) to determine $n(x)$ and ultimately $\phi(x)$. Figure 2.22 illustrates a finite fringe interferogram applying these principles.

2.4.4 High Sensitivity Interferometers

When the interest field is an extremely weak phase modulator, steps must be taken to either increase $\phi(x)$ or to increase the accuracy of measuring it. For example, to accomplish the former, the wavefront can be passed back and forth through the field many times. A Fabry-Perot interferometer becomes a powerful detector of phase change, when the phase object is placed in the space between the reflectors. In fact, a laser itself is perhaps the most sensitive detector of phase shift of all interferometers when the disturbance is placed in the laser cavity. The output power of the laser is extremely sensitive to phase changes within the cavity. This method has not seen significant application, however.

A more straightforward method increases the accuracy of the measurement of $A\phi(x)$.\textsuperscript{14-16} One such method measures directly the intensity of the mixed reference and test waves (Figure 2.23). We return to Equation 2.8 which has the intensity modulation term

\[
I_\text{m}(x) = 2u_0u_0 \cos [A\phi(x) + A\phi_0] \quad (2.14)
\]

If we limit the equation to small $A\phi(x)$ and set $A\phi_0 = \pi/2$, Equation 2.14 can be expanded to give

\[
I_\text{m}(x) = 2u_0u_0 \left[ 1 - k A\phi(x) \right] \quad (2.15)
\]

and the measured intensity is proportional to the phase shift $A\phi$. Physically this means that the output signal is nulled before the interest field is introduced. Such an interferometer is orders-of-magnitude more sensitive than those described in the previous sections. The interferometer must either be traversed across the field or other beam pairs introduced to define the entire function $A\phi(x)$.
Simple measurement of the intensity will not ordinarily be satisfactory in practice. This would require the laser to be noise free and single frequency. The problem of laser noise and beating can be solved by detecting the signal with what is known as a balanced detector.

Smeets and George have developed this instrument to a highly refined state and in the above reference they describe 14 different optical arrangements for dealing with practical problems. Figure 2.23b illustrates one such arrangement. A laser beam is passed through a quarter wave plate which shifts, say, the vertical component of polarization by $\pi/2$. The beam is split by a Wollaston prism into two normally polarized components which are collimated by a lens. They pass through a test cell, one passing through a phase disturbance, $\phi$. The two components are again combined by a Wollaston prism after which they can be expressed in the following way:

Vertical

$$\exp[i(\omega t + \phi + \pi/2)]$$  \hspace{1cm} (2.16)

Horizontal

$$\exp[i(\omega t)]$$  \hspace{1cm} (2.17)

They are then passed into a third Wollaston prism which is tilted such that its vertical direction is 45 degrees with respect to the first two. This separates the single beam of light into two components which have a combination of the above two components. The upper component can be written as the sum

$$\exp[i(\omega t + \phi + \pi/2)] + \exp[i \omega t]$$  \hspace{1cm} (2.18)

The intensity at the photosensor is, therefore,

$$2[1 + \sin \phi]$$  \hspace{1cm} (2.19)

The lower component can be written as the sum

$$\exp[i(\omega t + \phi - \pi/2)] + \exp[i \omega t]$$  \hspace{1cm} (2.20)

with the resulting intensity

$$2[1 - \sin \phi]$$  \hspace{1cm} (2.21)

The configuration shown subtracts the two intensity signals providing an electrical signal.
This method of signal detection is known as balanced mixing. It removes the bias term from the intensity, tends to cancel out constant noise terms and, thus, improves the accuracy and sensitivity of the measurement. It is the combination of the 1/4 plate with the last Wollaston prism which produces the opposite signs before the two sine terms above and provides for the bias cancellation.

One variation of this interferometer has been used to detect extremely weak phase changes in boundary layers with sensitivities for pathlength changes as small as 0.01 angstroms (Figure 2.23c). This allows the observation of subsonic flow disturbances, sound waves, and transition from laminar to turbulent flow. In the method shown, the signal is derived from light reflected from a surface, an ideal condition for observing flow over airfoils.

2.4.5 Phase Shift Interferometry

Phase shift interferometry is a relatively new technique that solves many of the problems in standard interferometry, especially the ones caused by coherent noise. Furthermore, the method is readily adaptable to microcomputers and new electro-optical components, therefore, simplifying data reduction. Writing Equation 1.8, which describes the Mach-Zehnder interferometer, in the following form

$$I_1(x) = I_a + I_b \cos (\Delta \phi(x) + \phi_0)$$

(2.23)

Let us consider that a controllable phase shift is added to the interferometer so that $\phi_0$ can be accurately set at will. There are three unknowns, $I_a$, $I_b$, and $\Delta \phi(x)$. These can be solved if three equations are produced. This can be done by setting three different values of $\phi_0$ (usually 0 and $\pm \pi/3$) and measuring the respective values of $I_1(x)$, $I_2(x)$, and $I_3(x)$. The solution of the equations for $\Delta \phi(x)$ is then

$$\Delta \phi(x) = \tan^{-1} \sqrt{3} (I_3 - I_2)/(2I_1 - I_2 - I_3)$$

(2.24)

The three interferograms can be presented to a CCD array which immediately extracts and stores the values of $I_1(x)$, $I_2(x)$, and $I_3(x)$, whereupon the computer can solve the equation for $\Delta \phi(x)$. This process is much faster and more forgiving of noise than locating fringe centers to determine $\Delta \phi(x)$. Also, it is no longer necessary to interpolate between fringes, a process that can involve significant error if the sampling is not properly done. Of equal importance is the automatic addition of sense to the data when the known values of $\phi_0$ are added. Normal fringe tracing methods in contrast to this method cannot determine if a fringe shift represents an increase or decrease in optical pathlength.

One drawback to the method is that three interferograms are required for each condition. This drawback can, however, be thought of as an asset in providing a way to average out noise by taking advantage of the controllability of the interferograms.

![Figure 2.23c](image)

Figure 2.23c High sensitivity laser interferometry for boundary layer transition detection. The upper Wollaston prism produces two separate beams that are focused to the model surface. Backscattered light from the two points is differentially phase shifted by turbulence in the region of the two points. The lower Wollaston prism recombines the beams into two beams that are adjusted to cancel each other by the feedback system.
A second requirement is that the intensities in Equation 2.24 be linearly related to the actual intensities in the interferogram. Therefore, an intermediate storage of the interferogram must remain linear. The only other specific requirement is that sampling be done at least once between each fringe so that the fringe count is not lost.

The method of phase shift interferometry has not been available long enough for it to have become a widely used method in aerodynamics; however, the apparent potential of this procedure is almost certain to lead to widespread use as soon as experimentalists fully exploit the method.

2.6.6 Heterodyne Interferometry

Another relatively new method is heterodyne interferometry, which is essentially the most general case for phase shift interferometry. In this case the fringes are continuously shifted in time while monitoring the intensity in the interferogram. We return to Equation 1.7, which describes the interference of two beams of different wavelength. In general, the interference pattern is a set of fringes that sweep over the field, such that at any point, the intensity varies at a frequency, \( \omega \), the frequency difference of the two interfering waves. The desired quantity for measurement is \( \phi(x) \), which is manifested as a phase shift on the time varying intensity at any point in the interferogram.

In heterodyne interferometry, the desired phase difference representing the unknown phase object is measured by monitoring the phase shift of the intensity signal sampled at each point in the interferogram by a small detector. An alternative to scanning a single detector is the use of a detector array. The only requirement is that the array density be high enough that at least one detector exists between each fringe in the interferogram.

There are a number of advantages in making the phase measurement in this manner. First, the accuracy with which phase can be determined is considerably improved over standard fringe tracing methods. Accuracy is limited by the ratio of the diameter of the detector to the fringe spacing and values better than one degree or wavelength/300 are commonly reported. Secondly, since the signal is a high-frequency electronic signal, it is more tractable to electronic filtering, noise rejection, and averaging than other methods. Finally, interpolation between fringes is not required since the signal exists everywhere in the interferogram.

Even though this method offers tremendous potential, it has not achieved widespread use in flow diagnostics at this writing. It seems evident that either this or the method of phase shift interferometry will eventually be important in flow diagnostics as the technology is developed and passed on to the users.

2.4.7 Local Reference Wave Interferometry

A primary difficulty in the use of interferometry is the stability requirement imposed on the system. In wind tunnel applications this is further complicated by the usual requirement of transmitting the reference wave over or under the wind tunnel, maintaining not only the stability but also eliminating the density gradients in the entire path of the reference wave. Local reference wave interferometry attempts to overcome this problem by extracting the reference wave from the data wave that has passed through the flow field.

To be interpretable quantitatively, an interferogram must have a reference wave that has a known phase distribution, preferably a uniform phase distribution such as planar. This can be done by taking a part of the data wave and spatially filtering all of the phase information from it. Figure 2.24 illustrates how this can be done. The method clearly works best when a large part of the flow field is undisturbed. When this is the case, the aperture should be so arranged to pass into the reference wave only the part of the data wave that is undisturbed.

The error in this type of interferometry can be estimated by examining the part of the interferogram from which the reference wave was extracted. If the fringes in that region are straight, then the error is small. If they are curving, then the error is of the order of the fringe variation.

Figure 2.24 Local reference wave interferometer. The reference wave must be spatially filtered by stops or other filters to remove all phase information.
If no part of the data wave is undisturbed then spatial filtering must be applied. One method of accomplishing this is with the commercially available, point diffraction interferometer. In this system the beam is focused to a point after having passed through the field of interest. At the focus, a mask scatters (diffractions) light from the centermost part of the beam to produce a reference wave while passing the remainder of the light as a data beam. The other means of performing the spatial filtering is to actually split the wave into two waves, one of which is then passed through a pinhole filter to serve as a reference wave.

Either method is extremely difficult to accomplish in practice without adding significant errors into the interferogram, errors that will be completely overlooked if the technique is not fully understood. (The interferogram will look perfectly normal). If the pinhole or diffraction mask is sufficiently small to produce an acceptable reference wave, the alignment is extremely critical and vibration sensitive, and the reference wave is extremely dim, requiring similar intensity reduction of the object wave. Therefore, the temptation is to increase the pinhole size, producing what looks like a better interferogram. The error in the local reference wave interferogram resulting from spatial filtering can be bounded by knowing the diameter of the spatial filter. A ray of light that misses the center of focus by 1.22 D/D (the diffraction limited focus radius, see Section 1.4.4) represents one wavelength of error in the original wavefront. Clearly, unless the reference wave is produced by filtering with a pinhole that is nearly the diffraction limited focus diameter, then ambiguities in excess of one wavelength can be anticipated in the resulting interferogram.

An exemplar wind tunnel application of local reference wave interferometry is the demonstration by Bachalo of such a system in producing high-speed interferometry movies of the flow over an airfoil in a transonic flow at the NASA Ames Research Center. The same principles have been applied in local reference wave holographic interferometry, described in Chapter 3.

2.4.8 Diffraction as a Type of Interferometry

Many important features of diffraction phenomena can be analyzed by applying principles of interferometry, and one may consider a diffracting edge to be, perhaps, the simplest of all interferometers. Consider the aperture of Figure 1.8. Light scattered from the edge has a phase difference with unscattered light. For the small angle case this difference along a normal-to-the-edge is given by Equation 1.22. The intensity maxima or fringes are defined by Equation 1.26.

Similarly, a circular aperture is an interferometer (see Figure 1.8). When the distance, $z$, along the optical axis is less than $r^2/\lambda$ (where $r$ is the aperture radius) intensity peaks occur along the axis each time

$$x_m = r^2 / \lambda \tag{2.25}$$

the separation between the maxima can be determined by

$$\Delta x_m = r^2 \lambda - r^2 / (m + 1) \lambda \tag{2.26}$$

At one far field ($m = 1$) this separation is

$$\Delta x_1 = r^2 / 2 \lambda \tag{2.27}$$

An interferometer of this type can be used as a vibration monitor or as a linear measuring device simply by monitoring the intensity scattered along the $z$-axis by a surface.

2.4.9 Reconstruction of the General Phase Object from Interferograms

An interferogram defines the phase of a wavefront after it has traversed the flow field along a given projected direction. In general, every different projection produces a different interferogram. The task is to reconstruct the phase object from the projections. This reduces to the definition of the so-called Radon transform. For radial symmetry conditions the Radon transform reduces to the Abel equation.

During recent years, related image reconstruction techniques have been successfully applied to several fields such as medicine, astronomy, electron microscopy, nuclear magnetic resonance, geophysics and optical interferometry. An extensive review can be found in Reference 21.

There are several difficulties associated with the reconstruction of phase objects from their projections:

(a) No sharp boundaries can be defined for phase objects, (with the exception of shock waves). The definition of the reconstruction region is based on a priori information about the flow field. In general, the contribution of the phase object outside the reconstruction zone must be assumed to be negligible.

(b) Relatively large errors are present in the interferometry data. The optical path differences are recorded as intensity of the fringe data ordinarily may not be better than 1/2 of the fringe spacing. In addition, for an unsymmetric density field, the number of fringes for each
projection varies for different view angles resulting in a nonuniform data set. The ideal projection data is a set of nonoverlapping, equally spaced, parallel rays covering the whole reconstruction region.

(c) Only the relative phase shift of the projection can be measured by the interferogram data. There is, therefore, a possibility of incorrect identification of fringe numbers. This results in a set of "slightly" inconsistent data.

The mathematical reconstruction method employed in such cases should not be overly sensitive to the "noisy" or inconsistent data.

Although the Radon transformation

$$\mathbf{r}(p, s) = \int \int n(x, y) \delta[p - x \cos(s) - y \sin(s)] \, dx \, dy$$  (2.28)

provides a rigorous solution to the problem of reconstruction from projections, the solution is uniquely determined only by an infinite set of perfect projections. Here, $\mathbf{r}$ is the optical path length, $n$ is the refractive index, $\delta$ is the Dirac delta function, $s$ is the projection angle, $p$ is the coordinate along the projection plane, and $x$ and $y$ are the coordinates describing the reconstruction region. In practical problems, the discrete nature of projection data and the unavoidable measurement errors may result in the failure of reconstruction. Most tomography codes are based upon some approximation of the application of the Radon transform with practical "tricks" for reducing the effects of such errors.

In this section, the iterative refinement method of least squares solution for tomographic interferometry developed by Tan, is presented. The effects of projection number, limited viewing angle and the measurement error on the reconstructed image have been determined for this solution through numerical experimentation. As an example application, the results of the reconstructed density field around the tip region of a revolving helicopter rotor blade that was reconstructed from 40 interferograms are presented.

Iterative Refinement Method of Least Square Solution

In principle, the ART (algebraic reconstruction technique) algorithms are the schemes for solving a large system of linear equations. The reconstruction region is divided into a square grid ($N = m \times m$ cells) and the refractive index within each cell or pixel, $n_{ij}$, is assumed constant. The Radon transformation then reduces to a set of discrete linear equations

$$\sum_{i=1}^{m} \sum_{j=1}^{n} W_{ij}(p, s) n_{ij} = L(p, s)$$  (2.29)

where $W_{ij}(p, s)$ are weight factors determined from geometric relations. It should be noted that only those factors associated with pixels through which the projection ray passes are nonzero. For $k$ different projection angles, each having $n$ nonoverlapping and equally spaced parallel rays, the number of equations $N$, is given by $N = k \times n$.

Equation (2.29) can be written in the form

$$\mathbf{C} \mathbf{X} + \mathbf{L} = \mathbf{0}$$  (2.30)

where, $\mathbf{L}$ and $\mathbf{X}$ are $N$ and $M$ dimensional vectors respectively, and $\mathbf{C}$ is the coefficient matrix with $N \times M$ elements. Equation (2.30) can be transformed into a symmetric, positive definite, normal equation given by

$$\mathbf{C}^T \mathbf{L} + \mathbf{C} \mathbf{X} = \mathbf{0}$$  (2.31)

In practice, due to measurement errors and other inconsistencies, Equation (2.30) becomes

$$\mathbf{C} \mathbf{X} + \mathbf{L} = \mathbf{R}$$  (2.32)

where $\mathbf{R}$ is the residual vector. An approximate solution is sought for which the Euclidean norm of the residual vector $\mathbf{R}$ is minimum. The Euclidean norm is defined as

$$\| \mathbf{R} \| = \sqrt{\mathbf{R}^T \mathbf{R}}$$  (2.33)

The solution of the normal equation gives the standard least square solution of Equation 2.30.

Procedures for obtaining solutions to a large system of linear equations can be found in a number of different scientific areas. Iterative techniques have been widely used. They begin with an initial estimate and repeatedly modify the estimate until some threshold condition is satisfied. There are different ways to modify the estimate. The effectiveness of a method depends on the matrix character. Considering the features of matrix $\mathbf{C}$ (large and sparse) and relatively large measurement errors, the so-called method of conjugate gradients is found to be applicable for image reconstruction.
The computer procedures of the conjugate gradient method have been detailed. Some of the results of applying the algorithm are presented here. This procedure runs on an IBM AT microcomputer, and is limited to 20 x 20 mesh size.

It can be proven that: The norm of the residual vector $R$ decreases with increasing iteration number. Thus, in principle, this code is a technique minimizing the residual $R$ to the level determined by the inherent errors associated with the computer truncation and measurement data. Therefore, the final reconstruction accuracy is determined by data noise. Here the errors associated with the measurement of the optical path length (fringe lines) are the major attribute to the data noise.

Tan has performed extensive numerical analyses to evaluate the convergence of the procedure to the correct values. This is accomplished by starting with a predetermined, exactly defined distribution producing synthesized interferograms. The study has evaluated the effect of errors in the measurement from the interferogram, the number of projections, and the total view angle. In this way, precisely known errors could be introduced numerically. From these studies the following can be concluded.

- The convergence rate is increased as the projection ray number, $N$, is increased. The corresponding inversion error for a sufficient number of iterations is, in general, independent of $N$.
- The norm of the residual vector, $R$, in all cases, unconditionally converges to a limit which is a strong function of the input data error. The resulting error for the first few iterations is independent of the input data error. However, for more accurate input data, additional iterations result in more accurate results. This underlines the importance of the accuracy of the data set. It also proposes a guideline for determining the maximum number of iterations based on the rate of change of the residual.
- Typically, a set of data can be inverted to convergence to the correct value within a few percent with five to ten iterations.
- Limited-look angles, ($\leq 30^\circ$) do not have a significant effect on the convergence rate or the accuracy of the reconstructed data. Acceptable results were obtained for a total view angle of $30^\circ$. Care must be taken to ensure that sufficient information is recorded on the interferograms when the view angle limitations are imposed.

An application of this code is presented in Chapter 3.2.2.4.
CHAPTER 3

AERODYNAMIC HOLOGRAPHY

3.0 BACKGROUND

A variety of good holography books are available and, therefore, the author makes no attempt to improve the already existing coverage of basic holography. However, for sake of completeness, a short treatment of the holographic process and its terminology is presented here. Also, a number of important developments that have not reached today's textbooks are incorporated here. This chapter emphasizes the application of holography in aerodynamics, especially the application to particle and flow diagnostics and the analysis and interpretation of holographic data.

Holography has satisfied a need that is so basic to the study of particle and flow fields that its use has continued to grow steadily since the first application. Even though the data handling problem has not been entirely solved, the number of applications requiring the features of holography has continued to fuel development and use. With present technology, holography is the only method for accurately recording a three-dimensional image of a dynamic event. This is particularly useful, for example, in the microscopic examination of combustion and explosive events, droplet breakup and formation mechanisms, and velocity measurement of droplet fields. Holography has likewise made some types of flow visualization possible that have not been achievable by other methods. Specifically, the applications of interferometry to directly compare two flow fields and applications of interferometry in facilities with poor optical windows are outstanding.

Holography was invented in 1947 by Dennis Gabor who produced holograms of a variety of microscopic samples using partially coherent light since lasers were not available for another fifteen years. His application exploited the magnification properties of holograms and not the three-dimensional imaging properties. Another twenty years passed before holography saw application in a field environment in the studies of fog droplets by Thompson and coworkers.

The first application to flow diagnostics was by Brooks et al., who used holographic interferometry to analyze the flow field around projectiles in flight. Since then, virtually every test facility in the world with a need for flow diagnostics has incorporated holography. One of the first major test facilities to incorporate holography into an operational wind tunnel was the USAF Arnold Engineering Development Center. By the early 1970s, systems were in operation in most of the other aerodynamic test facilities of the NATO countries.

The development of data handling techniques lagged far behind what was required to compete with conventional photographic techniques. In recent years, great strides have been made in removing this obstruction. Available computer power, new codes, and detector arrays have played a key role.

Flow visualization holography has not replaced more conventional methods, but rather has become a complement. When a distinct set of conditions or requirements exist, flow visualization holography is often the best or least expensive way, or quite often, the only way to obtain a desired set of data. It is otherwise usually, but not always, the more difficult method since it involves less refined, more complicated hardware.

3.1 Introduction to Holography

3.1.1 How Holograms Work

Ideally, a hologram of a scene can be thought of as a window into the past, through which a precise image of a scene can be viewed even after the actual scene is physically removed. Having the hologram is, from a viewing standpoint, equivalent to having an actual scene behind the hologram frozen in time. The radiation field on the viewers side of the hologram has the same information content in either case. This radiation field, which is unique to a given object field, can be extremely complex. It is synthesized by illuminating the hologram with a simple wavefront, which, through the process of diffraction, is converted into the more complex one. We are referring to the coherent radiation field which originally was reflected from or passed through the object field. Radiation generated by the object field must be excluded, since it is not coherent and it simply adds a bias to the coherent radiation; it is not involved holographically.

Holograms can be made by many methods on many types of material. They can be generated by computer, drawn by hand, or produced by photographic methods. The discussion here is limited to optical holograms. A hologram can be made by recording the intensity of the sum of two or more waves, one of which must be a reproducible wave (the reference wave), and the other of which is the modulated wave carrying phase and amplitude information (the object wave).

Holography will be explained both as a diffraction process in general and as a more simple reflection process for some cases. A hologram is a complex diffraction grating that diffracts the light of a reconstructing wave to form an image of the object field.
Figure 3.1 illustrates the formation and reconstruction of a hologram of a point object formed with a collimated reference wave. Note that as the angle between the object and reference wave becomes larger, the spatial frequency of the interference fringes likewise increases (see Chapter 2). When the developed recording is reilluminated, the wave is diffracted more by the higher frequency diffraction grating. The diffracted rays of light converge on a point to the right of the recording, and they also diverge from a point on the left side of the recording. The latter is called the virtual image, while the former is called a real image and is said to be the conjugate image of the latter. Because diffraction gratings exhibit this dual directional character, holograms always possess two images, one of which is the conjugate of the other.

Figure 3.1b illustrates this process in terms of reflection. Recalling that fringe planes form as a bisector of two interfering waves, with a spatial frequency that increases with angle between the two waves, an emulsion will be exposed to light sheets as shown. A processed silver halide emulsion will contain deposited silver layers in these regions that will act as mirrors to a reconstructing wave, diverging the light from a point where the original object point was located. This does not explain the simultaneous reconstruction of the conjugate point; however, a real image can be observed to occur from this process if the hologram is illuminated from its opposite side.

**Figure 3.1a** Diffraction Model. Note that the fringe frequency increases as the angle between the two interfering beams increases in forming the hologram. In reconstruction, the higher frequency grating diffracts the light through larger angles.

**Figure 3.1b** Geometrical Optics Model. The fringe planes create planar deposits of silver in the emulsion that act like tiny mirrors to reflect the light at different angles during reconstruction. Note that the fringe planes bisect the angle between the object and reference waves.

Figure 3.1 Holography of a Point
Figure 3.2 defines a general geometry for forming and reconstructing a hologram of a point object. A first order analysis for thin holograms, in which the angles between object and reference wave are small, results in the following set of equations for locating the reconstructed image:

\[
\begin{align*}
  x_1 &= \left(1 - \frac{1}{\lambda_2} \right) x_0 + \frac{1}{\lambda_1} x_p \\
  y_1 &= \left(1 - \frac{1}{\lambda_2} \right) y_0 + \frac{1}{\lambda_1} y_p
\end{align*}
\]  

(3.1)

These equations can also be used to determine the magnification in the reconstructed image.

When the object and reference waves are collimated, the holography equations become simply the relationships between angles.

\[
\sin \theta_0 = \sin \theta_r + \sin \theta_x + \sin \theta_y
\]  

(3.2)

Figure 3.2 Geometry for the formation and reconstruction of a hologram.

3.1.2 Classifying Holograms

Holograms are classified according to the recording material placing relative to the object and reference wave, the direction of the lighting of the object field, the type of lighting, and the use of lenses in the recording. Figure 3.3 summarizes this taxonomy. Two basic types of holography are in-line (made with the object wave almost parallel to the reference wave) and off-axis (made with a relatively large angle between the object and reference wave), as illustrated in Figure 3.4.

An in-line hologram can be made by passing a simple wavefront (plane or spherical) through the field, which must be about 80 percent transparent, and recording the interference pattern of the scattered (object) and unscattered (reference) waves. Off-axis holograms are made by mixing the scattered light with a mutually coherent wave that has taken a path around the field. Off-axis holography is much more versatile with respect to the type of object illumination.

Figure 3.3 Hologram Classification. Shown here are twenty-seven different classifications of a hologram.
An object field can be front lit or back lit, with direct, diffuse, or structured light. The most common type of structured light is the so-called sheet of light produced with a cylindrical lens and a small diameter collimated beam of light. Direct lighting is derived by expanding or diverging light directly from the laser.

Imaging with lenses or mirrors before recording is commonly done to impart a convenient size, location, or magnification to the field of interest, or to relax the required recording capabilities of the hologram. This is discussed in more detail later.

Other classifications describe the process of diffraction, the character of the recording material, and how the information is coded before recording. The diffraction process can be caused by amplitude transmission variations (amplitude hologram), phase shifting (phase hologram), or by path length difference on reflection (surface relief hologram). The process can occur in approximately a single plane (thin hologram) or in a volume (thick or volume hologram), and it can occur on transmission (transmission hologram) or on reflection (reflection hologram). More than one hologram can be stored on a single plate (multiplexing), or they can be stored in narrow strips by coding with cylindrical optics (integral holograms) or coded by filtering one of the dimensional dependences from the object wave (rainbow holography).

Clearly, a large number of possibilities exist. All of these have found use in some form in aerodynamic applications and there are many variations yet to be tried.

3.1.3 When to Use Holography

When the evaluation of diagnostic methods for a particular problem includes holographic techniques as candidates, one should ask if holography offers more than other recording methods. Holography is essentially an information buffer memory between the experiment and the data analysis. Holographic data storage often makes a standard technique applicable that would not otherwise be; when conventional imagery cannot store sufficient optical information, holograms sometimes can.

How many photographs are stored in one hologram? A resolution element, R, can be stored in a photograph if the element lies within a field depth $R^2/2
$, near a single focused plane (see Figure 3.5). A hologram with effective diameter D can resolve R if the element lies within a distance of $RD/R$, of the hologram. The ratio of these two is the number of photographs in the hologram, $N_p$, clearly a very large number.

$$N_p = 2D/R$$

EXAMPLE: A 10 cm diameter hologram is produced to cover a field of particles having 5 micrometer minimum diameter. This hologram could contain up to 40,000 different, useful photographs.

Holography is especially helpful in dynamic cases where it is required to record high resolution optical information over a large volume in a short time. But, in stationary cases, holograms may not necessarily be superior in storing the 3-D data, depending on how it is to be processed later. When a wavefront is to be analyzed for phase information, holographic storage is convenient since the hologram retains all optical information including phase. (But holography is not the only way to store phase information.) When the wavefront is extremely complex, such as that from a diffuser or from very poor optics, holography is likely to be the best if not the only candidate.
When a variety of wavefront diagnostics is desired, holography provides a method to store the wavefront until the diagnostics stage takes place. This is especially true when the experiment time is either short or extremely expensive, since data reduction can be done at a separate time and place at the scientist's convenience.

Finally, when ambient light levels are too great for photography, holography can be used to extend the recording range. The hologram acts like a coherent light filter. While ambient light does fog the hologram, it plays no direct part in the reconstructed wave. The reference wave intensity can be raised to greatly exceed ambient light level at the hologram. When mixed with the object wave coherently, it provides a type of amplification.

3.1.4 Hologram Quality

The quality of a hologram can be defined by three parameters: 1) diffraction efficiency, 2) resolution or spatial bandwidth, and 3) signal-to-noise power ratio or smallest retrievable object intensity. Diffraction efficiency depends upon the holographic fringe visibility, which is maximum when the object and reference waves are of equal intensity. (This is not where S/N is largest.) In some materials, diffraction efficiencies are above 90 percent, resulting in extremely bright holograms. Brightness is an important parameter in holographic gratings and in display, but it is of little importance in diagnostics.

Resolution is determined by experimental geometry, hologram size, and quality of optical components. The latter requirement can be replaced by a capability to accurately reproduce the reference wave used to make the hologram, as well as to reproduce the optical geometry in which the recording was made (to subtract out aberrations during reconstruction).

Signal-to-noise ratio (S/N) is determined by the type of emulsion and by the quality of technique such as cleanliness, linear recording, using liquid gates, and refined processing methods. Table 3.1 summarizes the techniques for improving resolution and signal-to-noise.

3.1.5 Recording Materials

Three types of recording materials are important in aerodynamic holography, silver halide emulsion, thermoplastics, and photochromics. The highest quality holograms are still made on silver halide emulsions, but the processing requirements of these materials greatly restrict their use. Thermoplastics and photochromics have been used only in very recent years, but they offer tremendous advantages to be described. Other types of recording materials that have seen little use in this field will not be discussed. A detailed discussion of recording materials is included in Smith's book, Holographic Recording Materials.14
Table 3.1

<table>
<thead>
<tr>
<th>WHERE APPLIED</th>
<th>SIGNAL-TO-NOISE RATIO</th>
<th>RESOLUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DESIGN</strong></td>
<td>1. Use as few optical elements as possible. Low scatter elements</td>
<td>1. Hologram size must be large enough</td>
</tr>
<tr>
<td></td>
<td>2. Baffle stray light</td>
<td>2. Emulsion resolution high</td>
</tr>
<tr>
<td></td>
<td>3. Spatial filter beams</td>
<td>3. Low object/reference beam angle</td>
</tr>
<tr>
<td></td>
<td>4. Choose an emulsion with low noise, high MTF, high Gamma</td>
<td>4. Use flat plates</td>
</tr>
<tr>
<td></td>
<td>5. Wave reference wave at least 10 times object wave</td>
<td>5. Place object close to hologram</td>
</tr>
<tr>
<td></td>
<td>6. Object reference angle is bisected by plate normal</td>
<td>6. Use direct (not diffuse) light</td>
</tr>
<tr>
<td><strong>DURING</strong></td>
<td>1. Keep everything clean in the optical train</td>
<td>7. High quality optics (flatness, parallelism, good lenses)</td>
</tr>
<tr>
<td><strong>RECORDING</strong></td>
<td>2. Maintain uniform intensity wavefronts</td>
<td>8. Image and magnify the object before recording</td>
</tr>
<tr>
<td></td>
<td>3. Put as little as possible in the object field</td>
<td>9. High coherence</td>
</tr>
<tr>
<td></td>
<td>4. Expose optimally</td>
<td></td>
</tr>
<tr>
<td><strong>DURING</strong></td>
<td>1. Develop to transmissivity of .16 - .20</td>
<td>Same as for high S/N</td>
</tr>
<tr>
<td><strong>PROCESSING</strong></td>
<td>2. Shorten development time</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Maintain cleanliness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Final distilled water wash</td>
<td></td>
</tr>
<tr>
<td><strong>DURING</strong></td>
<td>1. Keep optical train free of dust</td>
<td>1. Orient hologram position precisely</td>
</tr>
<tr>
<td><strong>PLAYBACK</strong></td>
<td>2. Uniform beams</td>
<td>2. Use real image corrected by projecting backwards through any optics originally used</td>
</tr>
<tr>
<td></td>
<td>3. Use a liquid gate</td>
<td>3. Use correct wavelength</td>
</tr>
<tr>
<td></td>
<td>4. Illuminate only the part of hologram required</td>
<td>4. Duplicate original reference wave</td>
</tr>
<tr>
<td></td>
<td>5. Spatial filter</td>
<td>5. Spatial filter</td>
</tr>
</tbody>
</table>

3.1.5.1 Silver Halide Emulsions - Figure 3.6 summarizes the properties of a typical photographic emulsion. The optical noise scattered from a beam of light incident upon the emulsion is greatest in the transmission direction, and the ratio of transmitted to scattered light begins to level off at about 30 degrees at a value of about 10−4. Note that the diffraction efficiency peaks at an intensity transmission coefficient of about 0.16, a number which has been found to apply to most photographic materials. The amount of noise scattered from the illuminating beam decreases with photographic density since it is absorbed in the emulsion. The diffraction efficiency peaks where the slope of the transmission versus exposure curve is maximum.

Figure 3.7 compares the characteristics of a variety of photographic materials. Note that the peak diffraction efficiency is about 3 percent. The fastest emulsion is Kokak SU-523, which is also the lowest in resolution. The most commonly used materials for ruby laser holography are Agfa 10875 and 8E75, while Kodak 120-02 is most widely used for HeNe laser holography. Kodak 649-F, first used for holography, produces excellent quality holograms but is much too slow to be of use generally.
Figure 3.6 Characteristics of silver halide emulsions.

Figure 3.7 Characteristics of common holography materials.

In producing the hologram, we are interested in four things related to the film:

1. maximum diffraction efficiency for a bright image,
2. high signal-to-noise ratio,
3. image resolution for high definition, and
4. minimum recording time.

The diffraction efficiency is simply the percentage of the reconstruction reference which is diffracted into the reconstructed object wave. The sources of noise include scattered light from the emulsion grain, surfaces of the emulsion backing and other surfaces, and contamination in the system. Parameters that affect desired hologram characteristics include the exposure of the hologram, its development, and the choice of a reference to object beam intensity ratio.

1. Exposure - For most films, the maximum diffraction efficiency occurs when the amplitude transmittance at the average exposure level lies between .4 and .5 for all object to reference wave beam intensity ratios. This corresponds to an intensity transmittance of .16 and a photographic density of approximately .8. The diffraction efficiency can easily vary by a factor of five over the hologram if the exposure of the hologram varies by a factor of five.

2. Development Time - A vast amount of control is possible during development. Normally, a shorter development time improves the signal-to-noise ratio, since the
diffraction efficiency increases slower than noise beyond normal development time. This means that it is better to overexpose the hologram in the first place and decrease development time to improve the signal-to-noise ratio. Also, this leads one to the conclusion that choosing a higher resolution film does not necessarily improve the signal-to-noise ratio if too long a development time is required. For example, one might otherwise choose Agfa 8E75 over Agfa 10E75, since the resolution of the former is higher. The conclusion here is that such a choice produces an improvement only if the development time remains short.

(3) Reference to Object Beam Intensity Ratio - The simplified analysis of the holography process usually makes the assumption that a linear relationship exists between the exposure energy and the amplitude transmission of the developed hologram. When such a relationship does not exist, the recording process is said to be nonlinear. Non-linearly recorded holograms produce not only the normal first order holographic image, but also higher order images. These images, in themselves, do not normally constitute a problem. However, further analysis shows that even first order holographic images in nonlinearly recorded holograms have added noise terms. Therefore, this results in a lower signal-to-noise ratio.

A wide linear range can occur in holography processes when the reference to object ratio is 10 to 1 or greater. When the reference to object ratio is less than this, careful development time is required to keep the recording linear. The brightest hologram is not necessarily the linear recording. In fact, the highest diffraction efficiency occurs when the reference to object ratio is equal to unity, so brightness is not the correct thing to optimize. Therefore, the hologram which has been optimized for resolution is normally not as bright as the hologram maximized for brightness.

Development time should be shortened to achieve the photographic density on the hologram of 0.8 by increasing the exposure during recording, and by carefully monitoring the hologram during processing. The normal way of processing a hologram is to develop and inspect the hologram occasionally during the development, using a suitable safety light. This practice can be made more accurate by comparing with a properly developed hologram placed side by side with the hologram under processing. A more scientific and less subjective way of making this decision employs an infrared light emitting diode and receiver inserted into the development tank. As the hologram develops, constant monitoring of the transmissivity can be made without affecting the hologram emulsion, which is not sensitive to the infrared radiation.

3.1.5.2 Thermoplastics - With the exception of lasers, the thermoplastic recording device (TPD) is probably the most important development for the field of applied holography. TPDs provide a method to record and develop holograms in place electronically. Figure 3.8 depicts the principle of operation. A transparent conductive film on a transparent substrate is coated with a photoconductive film and a thermoplastic layer on top.

**CHARGING**

- THERMOPLASTIC PHOTOCONDUCTOR

---

A UNIFORM CHARGE AS DEPOSITED ON THE THERMOPLASTIC PHOTOCONDUCTOR.

EXPOSURE LIGHT LIGHT

---

EXPOSURE TO LASER FRINGE PATTERN REGISTERS THE CHARGES THROUGH THE PHOTOCONDUCTORS.

**SECOND CHARGING**

+++ +++

---

RECHARGE INCREASES THE ELECTRIC FIELD ACROSS THE EXPOSED AREA OF THE THERMOPLASTIC.

DEVELOPMENT

---

HEATING OF THERMOPLASTIC CAUSES PERMANENT DEFORMATION. THE IRREGULARITIES WILL DEFRACT LIGHT TO RECREATE THE ORIGINAL IMAGE.

ERASURE

---

ERASURE IS DONE BY CONTROLLED HEATING.

RECORDING-ERASURE CYCLE OF THE THERMOPLASTIC HOLOGRAM.

Figure 3.8 Thermoplastic Recording Principle of Operation
To make a recording, a positive charge layer is added to the thermoplastic. Exposure to light redistributes the charge so that the capacitive forces compressing the thermoplastic are greatest where the exposure occurred. A second charging enhances this effect even more. Heating the thermoplastic allows it to flow under the electrical forces. Finally, cooling the thermoplastic causes it to set permanently in its deformed state, resulting in a phase hologram. The entire process takes a few seconds.

Figure 3.9 shows the characteristics of typical TPDs. The sensitivity of this material is similar to that of Agfa BES5 emulsion; however, the resolution is, in general, worse than that of silver halide emulsions. The diffraction efficiency is considerably higher since a phase process is involved. Since the material is insensitive before charging and after development, it is easier to use.

![Figure 3.9 Characteristics of typical thermoplastics](image)

TPDs suffer some disadvantages, however. A high voltage process is involved in the charging of the plate, the materials are sensitive to moisture, dust, and temperature, and the initial cost of a system is relatively expensive. The available formats are limited to 35 mm and 70 mm. Probably the most serious limitation is the relatively low signal-to-noise ratio compared to most photographic materials.

Two system types are commercially available. One of these is a glass backed erasable material that allows the recording, erasing, and reusing of the plate up to about 1,000 times. This can be extremely convenient and cost effective, if the required data can be extracted from the hologram shortly after recording. The second type has a flexible backing and is not erasable. This is convenient and cost effective if each hologram is to be saved.

3.1.6 Holocameras and Reconstruction Systems

3.1.6.1 General Requirements - A holocamera is a system of lenses, mirrors, lasers, and other optics for making holograms. In aerodynamic holography, holocameras are specially designed to meet the requirement. Choosing from the array of possibilities described in Section 3.1.2, the system must be engineered to accomplish the following tasks:

1. Properly illuminate the field of interest.
2. Place the interest field or its image in a suitable place relative to the hologram.
3. Achieve an acceptable F-number for the desired resolution.
4. Deal with environmental factors.
5. Meet coherence requirements. This is done by splitting the laser beam into two beams, using one as a reference wave and the other as an object wave, and bringing the same rays back together as accurately as possible after having traveled the same distance.
6. House the film or plate transport, laser, and optics.
7. Produce a hologram of acceptable quality.

The reconstruction system is the system of optics that provides for the replay of the hologram into a data reduction system. It may be part of the holocamera or it may be a completely separate system. Ideally, the hologram should be illuminated with the same wave as the original reference wave, but practical considerations often prevent this. So, when a hologram is made with a pulsed laser, a CW laser is commonly used in the reconstruction system, and it is not always possible to match wavelengths.
Changing the laser wavelength, angle of illumination, and radius of curvature for hologram reconstruction introduces aberrations into the final image that must be taken into account. Such aberrations are similar to the Seidel and chromatic aberrations present in lenses. A hologram that is illuminated at the wrong angle will exhibit astigmatism and coma, while one that is illuminated with the wrong radius of curvature exhibits spherical aberration. An aberrated reference wave translates with a linear transfer function directly into the reconstructed image.

A reconstruction system should incorporate the following elements:

1. Optics to produce a reference wave like the one with which the hologram was made.
2. Mechanical devices to position the hologram in the reference wave as it was originally recorded. These should be adjustable over few degrees.
3. A scanning system to scan through the three-dimensional image.
4. Controls that allow the viewer to scan the hologram while examining the image.
5. A recorder, such as a closed circuit TV and VCR or a photographic camera (or both).

Some reconstruction systems require special features such as the following:

6. Adjustments that allow the positioning and location of the hologram to wavelength accuracy.
7. Fixtures to hold and position two or more holograms relative to each other.
8. Allowance for more than one reference wave.
9. Independent phase and intensity control on one or both of the reference waves.
10. Computer controlled traversing and other mechanical adjustments.
11. Features that allow the reconstruction of the conjugate wave backwards through the original recording optics for the purpose of aberration removal.

The reconstruction system must be compatible with the holocamera and must be designed to produce an image that is sufficiently bright, noise free, aberration free, the right size, and in the right location for analysis. An extremely important function of this system is to cast the data into a useful form. The images may be recorded as photographs, or analyzed directly by an image analyzer. So the quality of the optics must be generally as good as the required image quality.

3.1.6.2 Typical Holocamera - Figure 3.10 illustrates a common holocamera design, using a pulsed ruby laser as a recording laser and a collinear HeNe CW laser as an alignment laser. The CW laser can also be used for reconstruction within the holocamera. This particular system uses a double-barreled laser to provide two separately Q-switched beams of light. (Most holocameras use only single beam lasers.) Note that the object and reference pathlengths are matched. The object field is imaged with a one to one imaging telescope to a position near the hologram. In a system like this where the same rays do not come back together at the hologram, it is important that the lasers have good spatial coherence. This type of holocamera is capable of producing threedimensional images of the object field with a resolution of about 10 micrometers. Multiple pulsing of the laser and use of both beams make possible the recording of multiple 3-D images, with time spacings from almost zero up to about 800 microseconds.

A ruby laser equipped system can make holograms typically with a 20 nanosecond exposure, which freezes the motion of most fields of interest. Other laser candidates for this type of system are YAG lasers and dye lasers. Ruby lasers were historically the first used, but YAG lasers are gradually becoming more competitive. Compared with the maximum one multiple pulse per second of commercial ruby lasers, YAG lasers can pulse at rates up to about 30 per second, making them ideal for making movies.

3.1.6.3 Viewing Reconstructed Images - A key element in reconstruction and analysis of holography data is the development of skill and technique in focusing coherent images. This can and should be developed upon images which are first generated directly without holograms, so that holographic effects can be understood separately. The hologocamera in Figure 3.10 has a built-in alignment laser that projects a continuous wave helium-neon beam of light through the system in the same space that the ruby laser will ultimately traverse. This alignment laser can be used for more than alignment, providing information about the reconstructed image, even before the hologram is recorded. This is done by placing a stationary object, in the field of view that will ultimately be occupied by the experimental object, and examining the image that is to be presented for holographic recording. This image can, by removing the frosted glass from the hologram holder, be projected into a working space which is directly accessible to the researcher. Therefore, many of the properties of the image can be studied in detail directly, without being confused by holography effects.

It is useful, especially for the novice, to practice focusing with this directly produced image using wires or sharp edges placed at the position of the object. This image can be examined in a variety of ways. It can be positioned directly on the face of a TV camera and examined on a CRT monitor. This is, perhaps, the most ideal method for
viewing, since no safety considerations need be taken. The image can be viewed on the frosted glass of a macro-camera simply by focusing the camera on the actual spatial image which appears behind the hologram position. Finally, for the experienced observer, the simplest method is to view the spatial image directly with a microscope or an eyepiece and no frosted glass. In this manner, the viewer can observe the actual image free of speckle, which is introduced when a diffuse surface is used. The image can also be examined in this mode with a macro-camera, if a clear focusing screen is used in the camera as opposed to the usual frosted screen.
EXAMPLE: When viewing these images directly, safety is a primary consideration. The maximum acceptable emission limit established by the American National Standards Institute shows that, under almost all circumstances, this direct image may be safely viewed with the proper CW laser. Acceptable Emissions, taken from Table 4 of the above guide, state that for the helium-neon wavelength for emission durations between 10 seconds and 10,000 seconds, $3.9 \times 10^{-2}$ joules of energy are permissible. A 4 milliwatt alignment laser, used in the holocamera described above, allows approximately 1,000 seconds viewing the direct laser beam at the hologram position before exceeding the maximum exposure limit. In the holocamera system, the beam reaching the hologram has been expanded from 2 mm to approximately 100 mm, reducing the laser intensity by a factor of approximately 2500.

To develop skill in focusing an image, one must understand a few characteristics of coherent light. Consider Figure 3.11. A sharp edge is imaged by lens into a space where it is examined in fine detail by a viewer. It is always best to search for a sharp edge or a small object to determine focusing effects most critically. The object is imaged as a shadow and in the exact image plane the shadow has very sharp edges with extremely fine diffraction effects.

![Diagram of focusing on a coherent image of an edge.](image)

Planes $A'$, $B'$, $C'$, and $D'$, are images of the light distributions in planes A, B, C, and D.

Figure 3.11 Focusing on a coherent image of an edge.

In the actual plane of the object, light is scattered by the edge both into the shadow region and into the region where light passes unscattered by the object. The scattered light and how it varies as the image goes in and out of focus can be observed with a magnifier in the image to determine with high precision the actual focus position of the image. In the illuminated region, this scattered light mixes with unscattered light, forming a diffraction pattern which at first is extremely fine. As one proceeds further and further out of focus, the diffraction lines become broad and more distinguishable.

One difficulty in focusing occurs when one is extremely out of focus. When the observed plane is badly out of focus, a motion toward the focus plane will not produce a drastic change in the diffraction pattern. Only when one nears the focus plane do the collapsing diffraction patterns and the illumination scattered into the shadow become clearly observable with rapid change such that an accurate focusing can be determined. Therefore, one must first find approximately the focus plane and then examine the microscopic details of the image edges.
A common mistake made by the novice is to expect that since parallel light is passing over an object, there is no focus plane. This would be true only if there was no diffraction. Diffraction allows one to locate, very precisely, the focus plane of the image. Similar focusing characteristics can be observed for wires, straight edges, and particles, and it is useful for an observer to practice focusing on such targets.

Focusing is best done on an object that is strongly diffracting, such as an edge. Phase objects also diffract light, but except for step changes in refractive index, they are more difficult to focus because they are not strongly diffracting. The effects described in the previous paragraph are not as easily observed. Therefore, when the field of interest contains no opaque objects, it is often difficult to locate exactly where the focused image is. This is another reason why it is important to locate the approximate position of the image, relative to the hologram, before the hologram is actually made.

3.1.6.4 Optimizing for Data Reduction - A principal advantage for holography is that the entire wavefront passing through the interest field can be recorded and reproduced exactly. In practice, however, real effects make it important to choose a configuration that has been optimized for the highest quality and easiest image analysis. Holograms can be made without lenses; however, it is also highly advantageous to use the best properties of holograms and lenses in combination to get the best results. Even so, every additional optical element added to a system increases optical noise. For this reason, one should limit to a minimum the number of lenses, windows, and mirrors required to produce optimal results.

The accuracy with which reconstruction angles are to be matched depends ultimately upon the desired resolution. Image resolution in the range from 5 to 10 microns does not require a highly critical alignment of the hologram. Image resolution below 5 microns requires that this angle be aligned very critically. This is commonly done by making the hologram of a resolution chart and adjusting the angular position of the hologram until the resolution is optimized in the reconstructed image.

It is usually helpful to position the image at a place that is not located in the overlapping reference wave, making it simpler to view without interference by the reference wave. It is also desirable to place this image at a location where it is easily accessible to a recording camera. At this stage, it may even be desirable to produce an image size that is the same size as the format of the camera. Again, this must be traded off with resolution. It is easiest to obtain high resolution by magnifying before holographic recording, thus making the overall reconstructed image size quite large. For this reason, large format cameras are desirable for reconstruction.

Further magnification of the reconstructed image is often useful, and is limited only by the quality of the hologram itself. The simplest high resolution image that can be obtained is the in-focus image in the plane of the hologram (known as image plane holography). This image, while actually reconstructed as a three-dimensional image, is (in the exact plane of the hologram) also a photograph and, therefore, is not as seriously affected by the holographic parameters as are images out of the film plane. This image can, be observed with incoherent light, and suffers least the type of aberrations such as chromatic, spherical, and astigmatism, commonly associated with holographic reconstructions. As a general rule, the further the distance from the hologram to the image plane, the more critical the hologram parameters (such as alignment) become.

3.1.6.5 Lens Configurations - In an imaging holcamera, such as the system described above, the possibilities for positioning and magnifying the image before and after reconstruction are many, and it is important to understand how these affect the ultimate appearance of the data. It is extremely important to know in advance the probable location of the image. This can save a vast amount of confusion and delay during the reconstruction. It also allows one to optimize the reconstructed data for whatever recording or viewing system is used. Three classes of imaging systems must be considered.

(1) A collimator pair which images one-to-one in three dimensions.

(2) A collimator pair which magnifies such that the magnification in the image field is uniform everywhere, but the magnification is not the same in the depth dimension as it is in the planar dimension. The lateral magnification is the ratio of the focal lengths of the two collimators used. The longitudinal magnification is the square of this ratio.

(3) Use of the single lens to image and magnify before reconstruction.

In the following discussion, each of these lens systems and its imaging properties is described in detail. For the first part of this discussion, the hologram itself is not mentioned. The reader must only remember that the hologram itself only recreates what exists optically, so, in any of these discussions, the hologram position could be at any place in the entire diagram. Ideally, with or without the hologram, the image space stays the same. One can consider the image space as being produced by the alignment laser and the hologram parameters when the hologram is not present, or being produced by the hologram when it is set up in a reconstruction system. From an optical point of view, when the optical wave from the hologram is examined, one cannot discern whether it has come from the hologram or from the imaging system in the holocamera, since the two should be identical.
Figure 3.12 describes the imaging properties of a collimator pair which images one-to-one in three dimensions. This system allows the largest field of view and, because the image space is identical to the object space, is the simplest system to analyze.

There are three separate cases that must be considered for use with a holocamera when the two matched lens collimator pair is used. Case I, which is illustrated in Figure 3.12a, has a sample space defined by a volume that is two focal lengths long and one lens diameter wide. The hologram can be positioned anywhere on the right-hand side of the second lens. In some cases where an image plane hologram is desirable, the image of the sample volume exists on both sides of the hologram, as shown in Figure 3.12b. As previously mentioned, it is sometimes convenient to position the hologram such that all of the sample space appears on only one side of the hologram. When the image exists on both sides of the hologram part of it is a real image and part of it is a virtual image. The real image can actually be presented directly to a frosted screen for viewing or directly to a lensless camera. The virtual part of the image must be reimaged into a real image before viewing on a frosted screen or recording.

Figure 3.12a One-to-one imaging in three dimensions with a matched collimator pair.

The part of the image that is virtual can be made real in holography simply by reversing the direction of the reference wave and projecting backwards through the hologram as shown in the third figure of Figure 3.12b. In this case the reference wave comes from the right and the viewer sees the reconstructed image on the left-hand side of the hologram. The part of the image that was virtual now becomes real. This process is often useful when presenting an image directly to the face of the vidicon, without using other lenses, where part of the field of interest happens to exist behind the hologram in a virtual space. This procedure allows one, by turning the hologram over, to convert this virtual space into a real space.

The two lens image methods just described applies both lenses before the recording of the hologram. In an equivalent technique, one can apply one of these lenses before a holographic recording and the other during holographic reconstruction. This is illustrated in Figure 3.12c. The recording is made with the hologram in a position just forward of where the second collimator would have been placed. The diverging wave is then recorded in the hologram. During reconstruction, the reconstructed wave will be that diverging wave. The collimator that was used during recording can then be used as an exact matched collimator to collect this diverging wave and recollimate it, producing
a one-to-one image of the original sample volume. The points A', B', and C' in this image correspond to the original A, B, and C in the sample volume.

Figure 3.12b Holography with a one-to-one imaging system.

To remove any imaging defects (see Figure 3.12c), the conjugate wave traveling antiparallel to the original wave is projected backwards through the lens used during recording, and the image appears once again in the original sample space.

Figure 3.13 illustrates the process of imaging with magnification. Two separate cases need to be considered—one in which the image space is a constant magnification of the object space, and the other in which the magnification of the object space depends upon the location in that space. The top diagram illustrates the imaging for the former case. Here, two collimators with different focal lengths are used, taking a collimated wave from the object space and reproducing a collimated wave in the image space. The lateral magnification (x,y) is the ratio of the two focal lengths. Magnification in the z dimension can be seen from this diagram to be the square of the magnification of the x and y dimensions. For this lens, therefore, magnification in the x and y dimensions is 3, while magnification in the z dimension is 9. For example, the distance BC is imaged nine times the original object dimension.

This type of lens configuration is extremely useful because the image space is magnified by a constant value independent of the position in the original object space. The volume lying from one focal length in front of the short lens up to its surface is imaged to the right-hand side of the large lens in a space that is convenient for producing holograms. As with the previous lens, the hologram can be positioned at any place in the image space. Precisely the same method for handling the reconstructed image for the matched collimator pair can be used here. If any part of the image is a virtual image in the reconstruction, it can be converted to a real image simply by reversing the direction of the reconstruction reference wave, a practice which is achieved normally by turning the hologram over. The stretching out of the image space in this reconstruction does make the analysis of the reconstructed image require more space, if the object itself has
considerable z dimension. When such is the case, it is often better to reconstruct the wave back through the lens pair to create an image space identical to the original object space.

CASE 2: RECORDING WITH ONE COLLIMATOR

RECONSTRUCTING ONE TO ONE WITH THE SECOND COLLIMATOR BEING SUPPLIED DURING RECONSTRUCTION.

REMOVING LENS ABERRATIONS DURING RECONSTRUCTION

Figure 3.12c Using one lens during recording and the same lens during reconstruction to achieve one-to-one imaging.

When an object field is magnified with a single lens, the amount of magnification depends upon the position of the object. The closer the object is to the focal point of the lens, the greater the magnification. When a very thin object field is used, little problem is created, since the amount of magnification is easy to determine simply by knowing where the object or image field is. The equations are provided in Figure 3.13. The magnification is simply the ratio of the image to object distance; or, when either the image or the object distance alone is known, one of the other two relationships for magnification can be used. For example, if only the image distance is known, as might be the case in observing a reconstructed image, the magnification is provided by the difference between the image distance and the focal length divided by the focal length.

When a larger sample volume is desired with magnification, it is sometimes convenient to produce a constant magnification image space. This can be done either during recording, as illustrated in the top figure of Figure 3.13, or during reconstruction as illustrated in Figure 3.14.

By studying Figures 3.12 through 3.14, it is easy to see how, by using different lens configurations, the image space of the hologram can be placed almost anywhere. Without some predetermination of this image space location, it is likely to wind up in an inconvenient space and require reimaging. Locating the image space can be done by using the relationships shown in these figures, or as described previously, by inserting the calibration object in the object space and observing the location of its image relative to the hologram.
Interferometry is the mixing of two mutually coherent waves to produce interference fringes that can be used to determine the phase distribution of one of the waves relative to the other. In holographic interferometry (HI), one or both of the waves is derived from a hologram.

3.1.7.1 Techniques - Basically, there are about seven different types of HI (see Figures 3.15 and 3.16). These include: (3.15A) interferometry with one nonreconstructed wavefront serving as the interferometry reference wave, (3.15B) real time holographic interferometry, (3.15D) "common path" interferometry (e.g., wave shearing), (3.16C) interferometry with two time separated wavefronts from a single plate (e.g., double exposure or double pulsed), (3.16A) interferometry of wavefronts from two holograms (e.g., double plate or sandwich), (3.16B) interferometry of wavefronts reconstructed with two reference waves. Heterodyne HI applies two reconstructing reference waves of different wavelength, producing moving fringes. Phase shift interferometry phase shifts the reconstructed waves in steps by independently phase shifting the reference waves.

3.1.7.2 Equations - In the following, we analyze the most general case, where the interfering waves are derived from two different holograms made at different times, separated by \( \Delta z \), moved laterally and tilted with respect to each other between exposures, and made with two different reference waves. The resulting general equation can then be specialized to describe all of the cases illustrated.
CASE 3: RECORDING WITH ONE MAGNIFIER

Lense Added To Provide Constant Magnification

NORMAL RECONSTRUCTION

RECONSTRUCTION BACK THROUGH THE ORIGINAL PATH

Figure 3.14 Holography with Premagnification.

The general equation will describe double plate HI. By setting \( \alpha z \) equal to zero, double exposure, double reference wave HI results. By setting the two reference waves equal, double exposure HI results. By allowing one of the reconstructed waves to emulate a real time wave, real time HI results.

In this computation, we are interested in the effect of the relative position of the holograms on the interferogram. We assume that a plane reference wave is used with an originally plane object wave that has been modulated by a phase factor \( e^{i \phi(x)} \), and recorded first on a Plate 1, then on a Plate 2 which is located behind Plate 1 (Figure 3.17). Then the two holograms are illuminated by a plane reference wave simultaneously reconstructing object waves from both plates. Let us fix the \( z \) position of Hologram 2 and assume that the position of Hologram 1 can move after the recording, and compute a one-dimensional \( (x) \) hologram.

We ultimately require the general equation of interference of the two waves \( U_0 \) and \( U_2 \) as a function of \( \alpha z \). The following are the object and reference wave amplitudes in the front plane of the respective holograms.

\[
U_{D2} = u_{D2} e^{ix} (x \sin \phi_{D2} + i x_{D2} + i s(x - \alpha z \tan \phi_{D2}) + i s_{D2}(x))
\]  
\[ (3.4) \]

\[
U_{22} = u_{22} e^{ix} (x \sin \phi_{22} + i x_{22} + i s(x - \alpha z \tan \phi_{22}) + i s_{22}(x))
\]  
\[ (3.5) \]
A. HOLOGRAPHIC INTERFEROMETRY WITH A SINGLE HOLOGRAM

Real Time Wave

Hologram

Interferogram

B. REAL TIME HOLOGRAPHIC INTERFEROMETRY

Hologram

Interferogram

C. HOLOGRAPHIC WAVE SHEARING INTERFEROMETRY

D. TWO REFERENCE WAVE SHEARING

Figure 3.15 Holographic interferometry with singly exposed holograms.

\[ U_{01} = u_{01} e^{i \frac{2\pi}{\lambda} (x \sin \theta_{01} - \alpha \cos \theta_{01}) + \phi_{02}} \]  

(3.6)

\[ U_{r1} = u_{r1} e^{i \frac{2\pi}{\lambda} (x \sin \theta_{r1} - \alpha \cos \theta_{r1}) + \phi_{r2}} \]  

(3.7)

where \( \phi_{01} \) and \( \phi_{r1} \) are phase modulations of each wave, \( \phi_{01} \) contains the flow field information plus any aberrations due to poor optics, while \( \phi_{r1} \) describes the effect of poor optics on the reference wave, and \( \theta(s) \) is the aberration introduced by Hologram 1 on \( U_2 \). To make the hologram, object and reference waves are added, giving amplitude

\[ U_{t1}(x,y) = U_{01} + U_{r1} \]  

(3.8)

with intensity

\[ U_{t1}U_{t1}^* = U_{01}^2 + U_{r1}^2 + U_{01}U_{r1}^* + U_{01}^*U_{r1} \]  

(3.9)

Now we assume a hologram which has an amplitude transmissivity which is linear in exposure.

\[ T_{hi} = U_{t1}U_{t1}^* \]  

(3.10)

During reconstruction, the reference wave is returned to the hologram and is modulated by \( T_{hi} \).
\[ U_{01} = U_{ri} T_{hi} e^{i \phi_{ri}} \]  

(3.11)

\[ U_{1} = U_{ri} + U_{ri}^* U_{0i} e^{i \phi_{ri}} + U_{0i} U_{ri}^* e^{i \phi_{ri}} \]

The second term contains the reconstructed wave that is of interest. Namely

\[ U_{01} = U_{ri}^* U_{ri} e^{i \phi_{ri}} \]  

(3.12)

We now examine the mixing of the two waves at the second hologram surface. We assume that the wave from the first hologram propagates with negligible diffraction over the distance \( az \). The amplitude of the reconstructed wave from the first hologram at the second hologram surface is (Figure 3.17) (assuming the same wavelength is used)

\[ U_{012} = U_{0i} e^{i \frac{2\pi}{\lambda} az \cos \theta_{0i} + \pi (x - az \tan \theta_{0i})} \]  

(3.13)

We now mix \( U_{012} \) and \( U_{02} \) and square to get the interferogram intensity.

\[ I_1 = (U_{012} + U_{02})(U_{012}^* + U_{02}^*) \]

\[ = (U_{012}^2 + U_{02}^2) + U_{012} U_{02}^* + U_{012}^* U_{02} \]

\[ = I_0 + 2 \Re (U_{012} U_{02}^*) \]  

(3.14)
Substituting 3.12 and 3.13 into 3.14 leads to

\[ I_1 = I_B + 2 \Re \left[ U_{r1}^* U_{r1}^* U_{r2} U_{r2}^* + e^{i \frac{2\pi}{\lambda} x} \cos \phi_{01} \right] \]

\[ x e^{-i \left( x - \Delta \sin \theta_1 \right)} e^{i \left( \phi_{r1} - \phi_{r2} \right)} \]

(3.15)

Get \( U_{01} \) and \( U_{r1} \) from Equations 3.4 - 3.7.

To include the effect of moving Hologram 1 in the \( x \) direction during reconstruction, note that the hologram transmissivity remains the same except for an \( x \) translation of \( \Delta x \). Namely

\[ T_{h1} = U_{r1}^* (x + \Delta x) U_{01} (x + \Delta x) \]

(3.16)

and this must be used in the above equation. And

\[ I_1 = I_B + 2 \Re \left[ U_{r1}^* U_{r1}^* U_{r2} U_{r2}^* + e^{i \frac{2\pi}{\lambda} x} \cos \phi_{01} \right] \]

(3.17)

times

\[ \cos \frac{2\pi}{\lambda} \left( x \sin \phi_{r1} - \Delta \sin \phi_{r1} - (x + \Delta x) \sin \phi_{r1} \right) \]

\[ + \Delta \cos \phi_{r1} \]

\[ + (x + \Delta x) \sin \phi_{01} - \Delta \cos \phi_{01} \]

\[ - x(\sin \phi_{r2} - \sin \phi_{r2}) \]

\[ - x(\sin \phi_{02} + \Delta \sin \phi_{01}) + \phi_{02} - \phi(x - \Delta \sin \phi_{02}) \]

\[ + \phi_{01}(x + \Delta x) + \phi_{02}(x) + \phi_{r1}(x) + \phi_{r2}(x) + \phi_{r1} - \phi_{r2} \]

\[ + \phi(x - \Delta \sin \phi_{01}) \]
\[ I_1 - I_B + I_s \cos \left[ \frac{2 \pi}{\lambda} \left( \sin \theta_1 - \sin \theta_2 \right) \sin \phi_0 \right] + \sin \phi_0 \sin \phi_2 + \alpha x \sin \phi_0 \sin \phi_1 \]
\[ + (\alpha x - \alpha' x') (\cos \phi_1 - \cos \phi_0) \]
\[ + \left[ \theta_0 (x + ax) - \theta_2 (x) \right] + \left[ \theta_1 (x) - \theta_2 (x) \right] + \left[ \theta_2 - \theta_2 \right] \]
\[ + \left[ \theta_0 (x - ax - \theta_0 (x)) - \theta_2 (x) \right] - \left[ \theta_1 (x - ax - \theta_0 (x)) - \theta_2 (x) \right] \]

Note that we have still retained two different reconstruction reference waves so that the
equation is general.

Now consider some special cases.

Case 1

(a) Single plate \((\alpha x = ax = 0)\)

(b) Same recording reference waves \((\phi_1 = \phi_2\) and \(\phi_1 = -\phi_2)\)

(c) Single reconstruction wave \((\phi_1 = \phi_2'\) and \(\phi_2 = -\phi_2')\)

(d) Same original object wave \((\phi_0 = \phi_2)\)

The resulting interferogram is described by

\[ I_1 - I_B + I_s \cos \left[ \alpha x (x) \right] \quad (3.19) \]

Note that all of the aberrating effects have cancelled out of the equation. Also note
that reference fringes (finite fringe interferogram) could have been added by changing
conditions between the two exposures, so that \(\phi_1 \neq \phi_2\) or \(\phi_0 \neq \phi_2\) or \(ax \neq 0\).

Equation 3.19 describes double pulsed holographic interferometry, and shows how it
depicts the phase change \(\alpha x (x)\) between the two exposures.

Case 2 - Double Reference Wave Holographic Interferometry (DRWHI)

Assume:

(a) Single plate \((\alpha x = ax = 0)\)

(b) Same original object wave \((\phi_0 = \phi_2)\)

The interferogram is described by

\[ I_1 - I_B + I_s \cos \left[ \frac{2 \pi}{\lambda} \left( \sin \theta_1 + \sin \theta_2 + \sin \theta_0 - \sin \theta_2 \right) \sin \phi_0 \right] + \sin \phi_0 \sin \phi_2 + \alpha x \sin \phi_0 \sin \phi_1 \]
\[ + \left[ \phi_0 (x + ax) - \phi_2 (x) \right] + \left[ \phi_1 (x) - \phi_2 (x) \right] + \left[ \phi_2 - \phi_2 \right] \]
\[ + \sin \left[ \frac{2 \pi}{\lambda} \left( \sin \theta_1 - \sin \theta_2 \right) \sin \phi_0 \right] \]
\[ + \left[ \phi_0 (x - ax - \phi_1 (x)) - \phi_2 (x) \right] - \left[ \phi_2 (x - ax - \phi_1 (x)) - \phi_2 (x) \right] \]

Equation 3.20 describes an extremely powerful technique. An interferogram
containing the data, which is \(\phi_0 (x)\), is produced, and it contains a completely
controllable set of reference fringes. The controls include:

(1) Spacing adjustment of the reference fringes during reconstruction by adjusting the
relative angles of the two reference waves.

(2) Orientation of the reference fringes.

(3) Phase shift of the reference fringes by adjusting relative phase of the reference
waves.

(4) Fringe contrast adjustment during reconstruction by relative intensity adjustment
of the reference waves.

Note that the phase errors in the object and reference waves carry through directly
into the final interferogram. However, if the same reference waves are used for
recording and reconstruction, the error cancels.

Case 3 - Double Plate Holographic Interferometry (DPHI)

Assume:

(a) The special case where identical reference waves are used for recording and
reconstruction

\(\phi_1 = \phi_2 = \phi_2' = \phi_2 = \phi_1 = -\phi_2\)
(b) The hologram normal bisects the angle between the object and reference wave

\( \phi_c - \phi_{r1} - \phi_{r2} \)

The resulting interferogram is described by

\[ I_1 = I_B + I_S \cos \left( \frac{2 \pi}{\lambda} \left( 2 \phi_c \sin \phi_{r1} + \phi_{r2} \right) \right) \]

where \( \delta \) is the difference in the phase error introduced on the object beam by passing through the first hologram from that introduced onto the reconstruction reference wave for the second. If the hologram does not change its thickness between recording and reconstruction, this can be zero. Also note that the reference fringe phase can be shifted by a relative lateral adjustment between the two holograms during reconstruction.

**Case 4 - Double Holographic Interferometry with Tilt in One Plate**

(a) Consider a slight tilting of one of the holograms during reconstruction, say hologram Number 2.

(b) Assume \( \delta x = \delta x' \) and \( \delta x = 0 \)

and for identical reference waves during recording and reconstruction

\[ (\phi_{r1} = \phi_{r2} = \phi'_{r1} \text{ and } \phi_{r1} = \phi_{r2} = \phi'_{r2}) \]

Since the second plate is tilted during reconstruction

\[ \phi'_{r2} = \phi'_{r1} + \delta \]

Let

\[ \phi'_{r2} = \phi'_{r1} + \delta \]

where \( \delta \) is the tilt angle (See Figure 3.18).

![Figure 3.18 Reconstruction with a film tilt error.](image)

Equation 3.18 becomes for small \( \delta \)

\[ I_1 = I_B + I_S \cos \left( \frac{2 \pi}{\lambda} \left( \delta \cos \phi_{r1} + \sin \phi_{r1} - \sin \phi_{r2} \right) \right) + \left( \cos \phi_{r2} - \cos \phi_{r1} \right) \delta + \delta \phi_0(x) \]

(3.23)

Now consider a case where \( \phi_{r1} = \phi_{r2} \) and \( \phi_{r1} = -\phi_{r1} \). Then

\[ I_1 = I_B + I_S \cos \left( \frac{2 \pi}{\lambda} \left( \delta \cos \phi_{r1} \right) + \delta \phi_0(x) \right) \]

(3.24)

Note that the phase shift across the interferogram is minimized with tilt when the object and reference waves are symmetrical about the hologram normal during recording (i.e., when \( \phi_{r1} = -\phi_{r1} \)).
Case 5 - Real Time Holographic Interferometry (RTHI)

The reconstructed wave, made at time zero, is mixed with a real time wave that has passed through the region of interest, providing an interferogram that depicts the phase difference of the real time wave and the time zero wave. Equation 3.18 can be applied directly with some analogy.

This is normally done by viewing the illuminated real time scene through the hologram. Assuming that the hologram is not moved after its original recording, Equation 3.18 reduces to

\[ I_1 - I_2 + I_{11} \cos \left( \frac{2\pi}{\lambda} \left( \sin \gamma - \sin \gamma_1 - \sin \gamma_2 + \sin \alpha \right) \right) \] (3.25)

Note that if the same reference wave is used, all aberrations cancel. Even the aberrations introduced on the waves by the hologram substrate are removed, since they are the same for both interfering waves. This important result allows high quality interferometry to be done through relatively low quality photographic film and glass plates without added error.

Case 6 - Single Exposure, Live Reference Wave HI

A degenerative case of real time HI is that in which an interferometer reference wave derived from a laser is mixed with a reconstructed data wave. If the reference wave is passed through the hologram, then Equation 3.25 applies. This method, which has been used to cancel otherwise troublesome film aberrations, has been termed the FACT (Film Aberration Correction Technique) method. The advantage is that single exposure holograms can be used for holographic interferometry with considerable fringe control available during data reduction.

Case 7 - Local Reference Wave HI (LRWHI)

Transferring the reference wave around the field of interest can sometimes present impossible requirements, for example, in large wind tunnels, in remote cases, and holography between two moving platforms (such as aircraft). For such cases, the reference wave can sometimes be transmitted through a part of the field of interest that does not add phase information to it, for example, through a laminar part of the flow field. Indeed, even if phase information is added to the reference wave, it can be filtered from the wave before it is mixed with the object wave, as described for local reference wave interferometry in Chapter 2.

In LRWHI, the object wave is split into two or more waves after having traversed the field of interest, one serving as the information wave, the others serving as the reference waves. It is usually best, at this point, to remove from the latter that part of the wave that has traversed the strong part of the flow field by aperturing, if possible. After this operation the wave may be a satisfactory reference wave. This can be determined by looking at the resulting interferogram which characterizes the region from which the reference wave was extracted.

If there is no weak part of the flow field, then it is necessary to spatial filter the phase information from the wave by passing it through a pinhole spatial filter to make it a useful reference wave. Just how much error this adds to the final determination of phase can be seen by examining Equation 3.18, which contains a term that represents the aberration of the reference wave.

Case 8 - Glint Holographic Interferometry (GHI)

A degenerative case of LRWHI occurs when some part of the object or object field contains a suitable point source (glint) that can be used as an object wave (a mirror can actually be attached to the object). Consider the holography of the surface of a fast moving object, such as a projectile that has a shiny part. Light from the region of the shiny part may serve as a reference wave.

Deriving the reference wave from the object wave may provide several benefits. Producing a hologram of the surface of the projectile may, under normal conditions, require extremely short exposures to insure that the holographic fringes remain stationary during the recording. However, if the reference wave is derived from the object itself, both the object and reference waves are shifted equally in phase by the motion of the object, greatly relaxing the recording conditions on stability.

Of perhaps even greater importance is the impression of the same phase modulation on the object and reference waves by the media between the object and the recording. This results in the cancellation of such information in the recording as though it were not there at all. This might have application in holography through a strong flow field, or in the holography of an object at long distance through a disturbing atmospheric condition. This principle has been used in the holography of burning droplets in rocket exhausts.17

3.2 Applications

In the following, a selected number of applications are presented to illustrate the use and state of the art of holography in the aerospace field. Holography is now being used in hundreds of laboratories. It is not attempted, however, to provide details and credits for all of these here. Examples have been contributed by the laboratories...
visited or otherwise contacted, and this does not necessarily mean that they represent the originators of the various techniques shown.

3.2.1 Particle Field Holography

The study of dynamic particle fields was the first real application of holography, and is still one of the most important scientific applications. Holograms can optically freeze a three-dimensional, high resolution image of a particle field that can be studied microscopically long after the actual particle field ceases to exist. Multiple exposure holography further allows the study of the dynamics of the field. There is currently no other way to perform such a three-dimensional study.

The field has been enriched by many new refinements and techniques since its first use in the early sixties. Included in the new technologies are the following:

1. Methods for aberrations corrections to provide diffraction limited imaging with low quality optics.
2. Methods for subtracting out optical noise normally present in holographic images.
3. Holographic movies.
4. Methods to automatically extract the data from holograms.
5. Multiple wavelength holography to encode exposures.
6. Multiple reference wave holography to provide two independent looks at a phase or particle field in the same hologram.
7. Thermoplastic recording for in-place electronic processing.
8. Many types of particle fields have been studied using by holography including the following:
   1. Ice, snow, and water droplets in the field of meteorology
   2. Pollution particulates and fibers
   3. Spray nozzle development and characterization
   4. Fiber and particle generation in, for example, the fiber insulation industry
   5. Particle formation and breakup
   6. Explosions
   7. Impact phenomena
   8. Two phase flow
   9. Combustion
   10. Crystal formation and growth
   11. Microbiology
   12. Cloud study from aircraft
   13. Flow diagnostics by particle tracking

3.2.1.1 Spray Nozzle Diagnostics - Figure 3.19 illustrates data taken from a doubly exposed hologram of a spray nozzle. Though the difficulty of presenting three-dimensional data in a two-dimensional format has not been overcome, this figure illustrates one technique of presenting such data. An overall shot of the field is recorded at a relatively low resolution to provide a global look at the subject field. Then microscopic higher resolution inserts are added to show specific details. As previously computed, such a hologram contains many such photographs. The illustrated recording is from a 10 cm diameter hologram. So there are about 40,000 different 10 cm diameter photographs available (with 5 micrometer resolution) in this hologram and many more at lower resolution.

3.2.1.2 Particle/Shock Wave Interaction - Figure 3.20 illustrates the study of the rebounding of particles through a shock wave after impact with a hypervelocity projectile. This is a classic case in which high resolution analysis of a particle is required at two different times in a dynamic situation and where the particle location in the volume is not known. A photographic solution would have required many cameras focused at different depths in the field of interest.

3.2.1.3 Holographic Subtraction - Figure 3.21 illustrates the use of holographic subtraction to improve the signal-to-noise ratio in a reconstructed image. In this example, the hologram is made with two reference waves, one for the first exposure and one for the second. The reconstructed images of each exposure are illustrated in a and b. In c the two images are reconstructed simultaneously. The optical noise is quite severe in all of these. If, however, during the reconstruction process, one of the reference waves is phase shifted by 180 degrees, then the backgrounds will subtract and the images will reverse from dark to bright. The improvement in S/N is evident.

3.2.1.4 Data Reduction - Particle field holography is still severely limited because of the difficulty in extracting the data from the hologram. Some image analyzers exist that augment data extraction, but, as of yet, no commercial instrument can handle completely...
DOUBLE EXPOSURE HOLOGRAPHY OF A CONICAL SPRAY

Figure 3.19 Double Pulsed Hologram of a Conical Spray. In the two photographs to the left, the camera is focused at different depths in the volume. Note how particles and ligaments in focus in one photograph are not in focus in the other.

Figure 3.20 Hologram and images of particles penetrating the bow shock.
the problem of distinguishing in and out of focus images and the severe coherent noise commonly associated with such images.

The most basic data reduction procedure is to scan the image volume while counting and sizing individual particles. Particle field images are usually observed on a closed circuit TV monitor or in photographs. In either case, it is important to understand that the depth of field is a function of particle size, so that the sample volume must be corrected for particle size when particles are counted in a single plane.

An alternative method is to examine the scattered light field from the hologram. It can be shown that the Fourier transform of a scattered light field can yield the size distribution. Using this principle, Ewan (1984) and Hess et al. have demonstrated that the size distributions of particle fields recorded in holograms can be extracted almost instantly.

The limitations up to now on holography data reduction, and the requirement to record at relatively slow rates, is somewhat confining. Other particle sizing instruments that can provide real time data at high rates excel for some requirements, while holography is considered to be the desired method to attain global looks and occasional full field examinations.

3.2.2 Flow Visualization Holography

When a light wave passes through a flow field, the optical pathlength is longer in regions of higher refractive index, so that the parts of the wave that pass through such regions are phase shifted, and the emerging wavefront is modulated with phase information that characterises the flow field. As shown in Chapter 1, the refractive index is directly related to density in gases so the gas density is sometimes derivable from a measurement of phase of the wavefront.

Shock waves are characterized by a discontinuity in refractive index, and they strongly diffract light. Therefore, they can be focused in 3-D like any object. Smaller refractive index gradients refract the light at smaller angles and cannot be observed in an ordinary image of the field; their effect on the phase of a wavefront must be observed with other methods to be described.
In flow visualization, since the object is usually a phase object (for example, a distribution of refractive index), off-axis holography is almost always used. A wavefront that has been modulated by the flow field can be recorded in its entirety for later analysis. The later analysis may comprise any of the conventional methods, such as shadowgraph, schlieren, deflectometry, Hartmann, moiré, or interferometry (see Chapter 2). Choice of the method can be postponed, and, indeed, all of the methods can be applied. Moreover, many variations for each method can be applied. The seven types of interferometry can be applied, the shadowgraph focus can be varied, the schlieren knife edge can be adjusted, and the deflection sensitivity can be varied.

A unique advantage of holography is that a wavefront modulated by the optical system only can be subtracted from one that has both optical system and flow field modulation, removing the effects of the optics. This can be extremely valuable in large systems, where cost of large diffraction limited windows and mirrors can be prohibitive. This principle can be carried further to subtract effects of a steady part of an object field, leaving an unsteady part which might otherwise be overshadowed by the steady part. Figure 3.22a illustrates a case in which a weak unsteady flow is overshadowed by a strong steady flow and can be seen clearly only after subtracting off the steady flow effects. Note the vortex street in the model wake, the unsteady shock wave, and the free stream turbulence in Figure 3.22b. One can zoom in on specific parts of the flow field with very high resolution as illustrated in Figure 3.23, while other parts of the flow field do not require such resolution for analysis.

Figure 3.22a Potential flow field.
3.2.2.1 State of the Art System - Figure 3.24 illustrates a conceptual YAG laser based, state of the art holographic flow visualization system. YAG lasers have become popular in this application because of their repetition rate, which is higher than ruby lasers, the first used in this application. Other pulsed laser candidates are dye lasers and Alexandrite, which all have pulse durations in the twenty nanosecond range. CW lasers can be used if the entire system is isolated from vibration, or if the laser can be properly shuttered to sufficiently short exposures. In general, the system would include two individually modulatable reference waves, a thermoplastic recording device, and an on-line image processor to analyze the interferograms. The methods illustrated in Figure 3.15 and 3.16 would then be applicable.

A good example of this type of system has been developed by Craig et al. and is in use in the NASA Ames Research Center Two Foot Transonic Wind Tunnel. The system employs a frequency doubled YAG laser that produces 20 millijoules of energy per pulse. Craig has recently demonstrated the on-line capability of the system, by employing a thermoplastic device that can reconstruct data directly into an automated fringe reduction system developed by Tan.

This basic type of system has been in general use in many wind tunnels since the early 1970s, but only in recent years has hardware and technology evolved to the point of on-line data capability. The applications described in the following sections are largely variations of the system shown in Figure 3.24.

3.2.2.2 Shock Wave/Boundary Layer Interaction - The interaction of shock waves with boundary layers is an extremely important process in aerodynamics. This process is also too complex to completely describe analytically. Therefore, experimental aerodynamics must provide quantitative results for the development and evaluation of theoretical models and computer codes. Holographic flow visualization has played an important role in providing such results.
Figure 3.23 Wake region illustrating high spatial resolution.

REFERENCE WAVE PATH

REFERENCE WAVE

FLOW FIELD

M1

M2

HOLOGRAM:
A. Newport Thermo-Plastic Recording Plate
B. Film Transport up to 30 Frames per Second
C. Auto Plate Changing

ON-LINE DATA REDUCTION

Figure 3.24 Conceptual state-of-the-art holographic flow visualization system.
Figure 3.25, from the work of Havener, in the study of shock wave/boundary layer interaction, illustrates various capabilities offered by holographic flow visualization. These also provide a good comparison of three different flow visualization techniques: shadowgraph (a), infinite fringe interferometry (b), and finite fringe interferometry (c). The interferograms were produced by the double plate methods, one hologram being of the empty wind tunnel and the other being of the Mach 3 flow field. Figure 3.25d provides a more detailed view of a shock wave/boundary layer interaction with an interpretation by Havener. The axisymmetric parts and the two-dimensional parts of the flow field can be analyzed quantitatively from these interferograms, but the three-dimensional parts cannot.

Figure 3.25a Shadowgraph - Flow visualization holography of a Mach 3 flow. One hologram provides an array of flow visualization types.

Figure 3.25b Infinite Fringe Interferometry - Flow visualization holography of a Mach 3 flow. One hologram provides an array of flow visualization types.
Figure 3.25c Finite Fringe Interferometry - Flow visualization holography of a Mach 3 flow. One hologram provides an array of flow visualization types.

Figure 3.25d Separation of a turbulent boundary layer from a flat plate due to a shock wave impinging on the boundary layer.
Figure 3.26, from the works of Bryanston-Cross, shows a double exposure holographic interferogram of a similar case for transonic flow. Included with the figure is a viscous numerical solution to the flow. Clearly, the flow field compares well in some places and not in others. Considerable understanding of both the flow mechanics and the experiment must be commanded to fully interpret these results.

Figure 3.26 Comparison of a transonic wedge profile with side wall boundary layer separation and a viscous numerical solution.

Figure 3.27, from the works of ONERA, is a real time holographic interferogram of a transonic channel flow, showing the density distribution in the flow at three different flow conditions. The pressure distribution in the channel and the Mach number can be determined by assuming two-dimensional flow and counting fringes. The Mach number, as determined from pressure gauges, agrees well with the values determined from HI. The hologram is first recorded with no flow, then it is returned to the original recording position and reilluminated with the reference wave. The reconstructed wave mixes with the real time wave to produce the interferogram. This system employs a 1.5 watt argon laser that is shuttered acousto-optically to provide exposures less than one millisecond. One advantage of this type of system is that it can be used to produce high speed movies of a flow. A disadvantage is that the apparatus must be highly stable.

3.2.2.3 Internal Flow - The flow in cascades, inside combustion engines, and in other confined regions presents a new range of problems, including timing and optical access. Getting the light in and out is no trivial problem. The installation of windows that survive and do not disturb the flow is a science in itself. Figure 3.28 illustrates the
work of Rolls-Royce\textsuperscript{77,78}, using double pulsed hologram interferometry of the flow between compressor blades in a turbomachine. The flow is made visible because the position changes between exposures, as shown in Figure 3.28. This type of holography is used to locate the shock wave position in 3-D. It cannot be used to quantify the density.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.27}
\caption{Diagnostics of the transonic flow in a channel.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.28a}
\caption{Method by which the holographic system was applied to the first stage compressor fan during a test.}
\end{figure}
Movement of a normal shock between two pulses with a 2msec expansion.

Figure 3.28b Movement of the compressor blade fan assembly between two holographic pulses.

Figure 3.28c Photographic reconstruction showing the shock structure at the tip of the compressor fan.

Oftentimes, the experiment must be simplified to fit the capability of the measurement to obtain meaningful quantitative data. The flow cascade has been constructed in a two-dimensional form for this purpose. Figure 3.29 illustrates a double exposure HI of such a case, included with a computation of the flow field for comparison.

Another such example is shown in Figure 3.30. This is the flow field inside an aerodynamic seal. This illustrates the appearance of vortices in an interferogram.

3.2.2.4 Three-Dimensional Flow - As described in Chapter 2, tomographic solutions now exist for solving 3-D flows if enough interferograms are made at the correct viewing angles through the flow. Only a few such systems are currently in operation; however, the number is likely to increase in the near future. A good example of such a system is in operation in the U.S. Army Aeromechanics Rotor Laboratory at Ames Research Center.
Figure 3.29 Holographic interferometry of transonic flow in cascades.

The system is a 60 cm diameter system of the type illustrated in Figure 3.24, using a pulsed ruby laser. The object beam passes over the tip of a rotor blade, recording the flow in the region around the tip, (Figure 3.31).

By traversing the system laterally, the view angle through the flow is varied automatically. The system must be fired by a signal that indicates that the blade is in the proper position. Holographic interferograms are typically recorded every five to ten degrees. For the views that contain large density gradients, more recordings are needed. Figure 3.32** is a sample of the data taken in this facility. Note the appearance of the shock wave at different angles of view. In some of the angles of view, the shock wave edge is not visible.

This type of system produces so much data that some type of automation in data reduction is imperative. Considerable efforts to this end have been successful in producing systems that can digitize and analyze data from an interferogram in a few seconds. These will be described in a later section.

The reconstructed density distribution ($\rho/\rho_0$), produced by the tomographic reconstruction code applied to 40 projections is shown in Figures 3.32c-d for the heights $z = 0.5, 1, 1.5$ and 2 inches (chord $C = 3$ inches; the leading edge of the blade $Y/C = -0.5$; aspect ratio = 13.7). The optimum number of iterations was 3 to 6 for the experimental data. The computation time on an IBM PC was approximately 3 minutes per iteration.

3.2.2.5 Other System Types - Some applications require special system design. Examples include the requirement to increase sensitivity, to reflect light from the model, or to look at very large flow fields. Figure 3.33 is a system that reflects light from the model or a panel behind it. If the surface can be treated with a retroreflective coating, then very large areas can be recorded. Recently, holographic optical elements have been devised for the treatment of surfaces for flow visualization.
Figure 3.30 Flow field in a pneumatic bearing.

Figure 3.34 is a system that is designed to achieve improved sensitivity. The object beam passes through the flow field more than once, adding modulation to the wave each time it passes through the object. The object wave with the correct number of passes is selected holographically by pathmatching. This unique coherence filter forms a hologram of only one of the many object waves that strikes it. The others simply tend to fog the film with light that is not mutually coherent with the reference wave. To compensate for the fogging, a critical part of the technique is to bleach the hologram before reconstructing the data. Figure 3.35 is a set of data taken with the system to illustrate this point. The object is a heated ball with natural convection around it. As the number of passes through the field doubles, then so does the fringe density in the interferogram for essentially the same flow field.
Figure 3.31 Rotating blade tip region and the coordinates.

Figure 3.32a Data from Helicopter.
3.2.2.6 Data Reduction - Figure 3.3f represents the flow of data in the holographic data reduction process. The first two steps, producing and reconstructing the holographic image, have been described above. The remaining steps are described here.

For years, publications reporting uses of holographic interferometry focused on feasibility demonstrations, and the actual data output of such research was quite limited. More recently, attention has been directed to the complete extraction of data from the holograms. Furthermore, recent laser and recording technology makes possible the recording of thousands or even tens of thousands of holograms, placing an even greater demand on automated data reduction.

The basic information contained in an interferogram is the distribution of the relative phase of two waves. This information can be extracted in essentially three different ways.

(1) Analysis of a fixed interference fringe set.
   (a) Locate the coordinates of the center of interference fringes.
   (b) Identify a reference point from which to start.
   (c) Identify and number fringes integrally starting with zero at the reference point and adding (subtracting) where the phase is known to have increased (decreased). Each succeeding fringe represents an increase or decrease in the phase difference of the two waves by $\pi$.
   (d) The dependence of intensity in the fringe pattern is sinusoidal and it is most practical to locate maxima or minima, although, in principle, any equal intensity locus between any two maxima represents an isophase difference contour.

(2) Analysis of a moving interference fringe set. (Heterodyne Interferometry)
   (a) The fringes are swept at a known rate (by methods already described), while monitoring the intensities at discrete points over the interference region.
   (b) Locate a reference point and arbitrarily define the phase as zero. As the fringes sweep over this point, a detector will produce a time varying sinusoidal signal.
Figure 3.32c: Reconstructed density fields for different heights above blade chord line.

Figure 3.32d: Density ratio profiles (blade length B = 47.1", chord length C = 3").

Figure 3.33: Large area flow visualization.
Figure 3.34 Multiple pass holographic flow visualization.

(a) SINGLE PASS

(b) DOUBLE PASS

(c) FOUR PASS

Figure 3.35 Multiple pass holographic interferometry.
Figure 3.36 Data flow in holography.

(c) Moving out from this point with a second detector, a sinusoid of the same frequency, but different phase, will be observed. The phase difference in the two time varying signals is the desired phase difference between the two wavefronts. Each time this phase difference increases through $2\pi$, the signal appears identical to that observed at the reference point.

(d) Numbering fringes becomes less of a problem, since the sign of phase shift is observable.

(e) Interpolating between fringes is no longer a problem, since they sweep over the detector. The resolution of the measurement is determined by the ratio of the detector size to fringe spacing.

(f) The entire process can be achieved with one fixed and one moving detector or, alternately, a fixed detector array.

(3) Analysis of the Fringe Set by introducing known phase shifts while observing intensity changes at discrete points. (Phase Shift Interferometry)

(a) A detector is positioned at a reference point where the phase is arbitrarily set at zero.

(b) Moving out from the reference point in steps of $\lambda$ less than $2\pi$, the intensity is given by Equation 3.18.

(c) The unknowns are $I_0$, $I_1$, and $\phi$. By introducing two known phase shifts, three equations are produced.

(d) The three equations can then be solved for $\phi$ in terms of $I_1$.

(e) A moving detector or detector array can scan the interferogram.

Unfortunately, the intensity in a real interferogram is much more complicated than that expressed in Equation 3.18, because of optical noise and parasitic interference. A real interferogram is characterized by:

(1) Broken, discontinuous, and split fringes

(2) Varying contrast

(3) Extraneous fringes

(4) Diffraction noise

(5) Broad cloud-like fringes

(6) Closely spaced fringes
Lack of a known reference position
Unknown fringe sign
Uneven background
Regions blocked by opaque objects
Speckle noise

Therefore, considerable preprocessing is sometimes required before the task of locating fringe position or measuring phase difference is practical.

Signal-to-Noise Improvement - In addition to noise in the interferogram, electronic noise in the sensors also contributes to the final results. One of the most direct methods of improving signal-to-noise ratio is signal averaging. Image analyzers can be programmed to average many data frames at TV rates, thus improving picture quality. Next, spatial smoothing is useful to reduce high frequency noise due, for example, to speckle-like noise.

Thresholding and Nonuniform Background - Many of the techniques commonly used in pattern and line recognition fail in interferometry because the intensity is sinusoidal and the background is nonuniform. This presents a problem in thresholding. Procedures for dealing with nonuniform background include the following: logarithmic detection, normalization and referencing to a mean intensity across the interferogram, which is first established before thresholding begins, normalizing to local intensity averages, and, finally, subtracting off the nonuniform background.

A floating threshold method, used by Becker, requires two passes over each line across the interferogram. The first pass searches for extrema. With the found extrema, a step-like threshold function is then defined, which is the mean gray value between adjacent extrema. The second scan is referenced to this function to produce a binary fringe pattern. This produces good results on fringe fields, even with severe background intensity variation and low contrast.

The problem can also be approached optically, in double reference wave holographic interferometry, by phase shifting one of the reconstructed waves by 180 degrees to subtract away the background entirely.

Interactive Preprocessing - Until artificial intelligence is applied to this problem, there is little hope of producing a fully automated system. Image processing systems fortunately allow a direct interaction from the operator to connect broken fringes, give the correct sign to fringes and to guess where a discontinuous fringe should proceed. What can be automated is partial recognition and flagging of the occurrence of such events to allow an operator to intervene.

Coordinate Extraction - After the above preprocessing, the fringe pattern is a binary pattern. Then this pattern can be reduced to a line pattern where lines are the centers of the black or white fringes. Alternately, the edges or transition from black to white can designate the fringes. Each scan line, therefore, provides a y coordinate (for the line), an x coordinate for the fringes. With each x coordinate, a fringe number is required. Each adjacent scan must then be referenced to the previous scan, so that like numbered fringes can be identified.

Because fringe density becomes very large in some regions, a magnification must be performed to maintain accuracy. This further complicates the numbering of fringes. A number of rules exist for fringe numbering.

(a) Adjacent fringes differ in number by zero or one.
(b) Fringes of different numbers cannot touch each other.
(c) Fringes cannot end inside the field of view (except where they close on themselves).
(d) When integrating over a closed line, the fringe number differences add up to zero.

Phase shift interferometry and heterodyne interferometry both offer promise for making automated data reduction of holograms easier yet. The output signals appear to be more amenable to filtering and noise removal. This is one of the most promising and rapidly growing areas in holographic interferometry today.
CHAPTER 4
LASER VELOCIMETRY

4.0 BACKGROUND

Laser velocimetry (or anemometry) has become one of the most popular laser diagnostic tools in the field of aerodynamics. Various methods are routinely employed in virtually all aerodynamics research facilities, universities, and companies using fluid mechanics. Organizations like ONERA (France), DFVLR (Germany), NASA (USA), ISL (Germany/France), RAE (United Kingdom), and the military services supporting aerodynamics have invested heavily in elaborate, computer controlled LV systems for use in wind tunnels and flow research facilities. These have undergone continuous evolution now for over twenty years.

The first principles of the methods and applications to simple flow fields are simple and straightforward; however, the general analysis and rigorous interpretation of the physics involved is extremely complicated, and a number of controversies still exist in the field. Among such controversies are the methodologies for proper seeding of the flow with scatter centers (SC) that accurately track the velocity fluctuations and the interpretation and removal of biasing errors.

Most laser velocimeters can be described as time-of-flight instruments. They function by measuring the time taken for a SC to traverse a structured and fully characterized amplitude or intensity distribution in space. The simplest structuring of light is, perhaps, the one in which two focused spots, with an accurately known spacing, constitute the sample volume (Figure 4.0) first suggested by Tanner\(^2\) and reduced to practice by Schodl\(^3\)\(-6\) and Smart\(^7\)\(-10\). The task is then to determine the time required for a scattering center to traverse the two points. These systems have been called two-spot systems, laser transit anemometers (LTA), and laser transit velocimeters.

A somewhat more complicated configuration exploiting the Doppler effect to determine velocity of the SC. As will be shown, this can also be described as a time-of-flight method. The laser Doppler velocimeter (LDV) or anemometer (LDA) can take on many different configurations. Most of these can be modeled by the so-called fringe model, which will be discussed in detail. The sample volume is structured by crossing two or more focused coherent beams of light. The crossover volume contains a set of well-defined interference fringes that can be used to track an SC as it traverses the volume. The set of fringes will be "real" or "virtual", depending on the configuration used. The basis for these systems was set forth in the works of Yeh and Cummins\(^1\)\(^1\) and was reduced to practice by many workers such as Rudd\(^12\), Durst and Whitlaw\(^1\)\(^3\), and Brayton\(^14\)\(-16\) and Mayo.\(^17\) Stevenson, et al\(^18\) have published a bibliography of early works.

4.1 Laser Doppler Methods

4.1.1 Theoretical Basis

4.1.1.1 Doppler Theory - When a coherent beam of light of frequency, \(f_0\), strikes a moving SC, an observer sees scattered light with frequency \(f_0 + f_d\), where \(f_d\) is the familiar Doppler frequency shift of the light. The shift occurs because the SC has a velocity component in the \(\mathbf{R}_d\) direction (Figure 4.1), causing it to receive wavefronts at a faster or slower rate than \(f_0\), and also because it has a velocity component with respect to the observer, causing either a compressing or spreading of the scattered wavefronts in the \(\mathbf{R}_d\) direction of the observer.
The observed phase of a wave, scattered by a stationary SC, will be changing at a rate \( \omega_0 \) (or \( 2 \pi f_0 \)) radians per second. However, the phase at a moving SC is seen to change at a rate \( \omega_0 - \frac{v \cdot \mathbf{k}}{c} \), and the observer sees still a different rate of change of phase given by \( \omega_0 - \frac{v \cdot \mathbf{k}}{c} \). The Doppler frequency is, therefore,

\[
\frac{d\phi}{dt} = \frac{v}{c} (\mathbf{k_s} \cdot \mathbf{v})
\]

(4.1)

Note that the Doppler frequency is proportional to the velocity component along the vector difference of the illumination and observing directions.

**EXAMPLE:** An SC traveling 100 cm/sec is illuminated normal to its velocity \((\lambda = 0)\) with an argon laser \((\lambda = 0.5 \text{ microns})\), and is observed at an angle, \( \theta \), of 60 degrees. The Doppler frequency is 100 cm/sec \(\times\) \(\cos 60/1\), which is about one MHz.

An alternative method is to illuminate the sample volume with two beams separated by some angle. The two beams are Doppler shifted by different amounts, and the detected Doppler difference can be used to determine the velocity component along the vector difference between the two illuminating directions. Equivalently, the scattered Doppler shifted radiation can be detected at two different angles and, as will be shown, this Doppler difference is proportional to the velocity component along the vector difference of the two directions. We will show how to analyze each of these cases by both a Doppler/heterodyne analysis and a fringe model analysis. The systems that collect the scattered light, and measure the amount of the Doppler shift with reference to the unshifted light are referred to as reference-scatter (RS) systems. Those that use two components of scattered and Doppler shifted light are called Doppler difference or dual-scatter (DS) systems.
The remaining problem is to measure the Doppler frequency. Optical frequencies lie in the range of $10^{14}$ to $10^{15}$ Hz and cannot be measured directly with electronic devices. Essentially, two ways are used: 1) interferometry, and 2) heterodyning (or "beating down" the frequency, by mixing a second wave of slightly different frequency with it to produce a lower frequency signal that can be analyzed electronically).

4.1.1.2 Fringe Models - In wave theory, the phenomena that involve frequency shifts can often be described not only with time domain equations, but equivalently with spatial domain, phase shift, or interferometry equations. The Doppler effect is such a phenomenon, and the so-called fringe models for LDV are examples of interferometric analyses of the Doppler effect. Fringe models have played an extremely important role in the evolution of LDV because they have simplified the understanding, design, and use of the technique. Consider the configuration of Figure 4.2, in which the SC is illuminated by two focused, crossed beams that have equal intensities and Gaussian profiles. Referring again to Figure 1.3, it will be recognized that the crossover region will contain a set of stationary fringe planes. Figure 4.3 summarizes the character of the fringes and the sample volume. As an SC passes through the fringe planes, it scatters light whose intensity is modulated at a frequency

$$f_d = (2 v_o / \lambda) \sin(\theta/2)$$

(4.2)

Figure 4.2 Dual scatter laser velocimeter, heterodyning at the scatter center.

Figure 4.3 Enlarged cross-sectional views of the beam crossover region.

The so-called geometrical sample volume, if determined by the transmitter optics, is represented by the volume wherein good fringes exist. Along the z-axis and in the xy plane, the fringe contrast is unity. For laser beams with Gaussian profiles, the optical intensity varies in such a manner that the fringe contrast in the sample volume is not unity anywhere else. A point SC scatters light uniformly in all directions, and it scatters proportionately to the intensity to which it is exposed. The form of the detected signal is plotted in Figure 4.4, for traces of a point SC traversing the sample volume in the xy plane.

The geometrical properties of the sample volume can easily be determined from Figure 4.3 when the transmitter is the defining system. The surface on which the intensity falls to $1/e^2$ of its maximum value is an ellipsoid.
EXAMPLE: & = 6 degrees, i = .5 microns, lens f/no f/2b = 200. Then the sample volume diameter is .013 cm by 25 cm long, and the frequency to velocity conversion factor is 2 kHz/cm/sec.

Figure 4.4 Intensity distribution in plane of crossed beams. (Assumes beams of equal intensity and polarization vectors linear and parallel).

The geometrical properties of the sampling volume can also be determined by the receiver. This is easiest to see when an off-axis detector is used. To determine the actual geometrical sampling volume, one can image the receiving aperture into the transmitter-formed sampling volume. The signal can come only from the overlapping region.

So far, we have referred to the geometrical sampling volume as being determined by the transmitter or receiver. The actual volume from which data can be taken is much more complicated than this. Consider a case in which SC of a single size are used. The volume from which data is taken depends upon the detectability of the scattered light. This is determined not only from the geometrical properties of the system, but also from the electronic properties of the detecting system, for example, by where the detection threshold is set or by other restrictive signal processing. The actual sample volume can be modified electronically by deliberately placing conditions on the signal. For example, the signal can be required to contain a set number of fringes above some threshold. One might also choose to accept only those signals having contrast greater than some value. Clearly, the actual sample volume depends upon the SC size for any geometrical configuration.

Consider the system of Figure 4.5. Here the sample volume is illuminated by a single beam, but light is collected at two different angles. No real fringes could be observed in such a sample volume. Nevertheless, it is useful to introduce the concept of virtual fringes, fringes that would exist if the two received beams of light were back-projected to form a sample volume. We will show that the signal from such a system is accurately described with this model for a wide range of system conditions.

An interesting reciprocity relationship exists between the laser and photodetector. The laser and photodetector position can be switched for any given system geometry resulting, in an alternative configuration with similar geometrical properties. Figure 4.2 and 4.5 are examples of reciprocal systems.

The parameters of an RS system can be determined to a large extent by using the same fringe model. The method is to examine a virtual sample volume that is formed by
the virtual crossover of the input beam and the scattered beam. The scattered vector changes within the sample volume, however, so that the Doppler frequency varies over the sample volume leading to the expression "Doppler Ambiguities". This volume will contain a set of virtual fringes of the proper spacing and orientation to evaluate the system. However, it is more accurate to examine fringes at the place where mixing effectively takes place. In some DS systems, and in all RS systems, this occurs at the photodetector. Therefore, the signal from such a system can be accurately modeled as an interference fringe that is formed by the mixing of a scattered, Doppler-shifted wave with a second reference wave producing a beat frequency. This model is useful in determining a variety of instrumental effects such as misalignment and poor quality optics.

4.1.1.3 The Moire Analogy - To augment the fringe model, Durst and Stevenson19 adapted the moire/interferometry analogy to LDV. By using the grids supplied in Appendices 1-4, many of the characteristics of LDV sample volumes and signals can be illustrated. Most of the effects that can be modeled with the fringe model can be demonstrated with the moire analogy, (see Chapter 2). Practical effects such as beam misalignment, vibration, frequency modulation of one of the beams, multiple SCs in the sample volume, wavefront distortion, and heterodyning of signals can all be lucidly illustrated. These and other effects are left as an exercise for the reader.

4.1.2 Measuring the Doppler Frequency

4.1.2.1 Direct Optical Frequency/Wavelength Measurement - As shown in Chapter 2, interferometry can be employed to determine optical wavelength and, thus, to measure Doppler shift. The first interferometer to be adapted to this measurement was a Fabry-Perot interferometer.24-27 Figure 4.6 illustrates the configuration. The interferometer is swept through its free spectral range by piezo-electric elements that vary the mirror spacing. The interferometer is illuminated by both Doppler-shifted and unshifted radiation, the latter serving as a reference. A velocity measurement is made each time the interferometer passes the Doppler-shifted light, typically up to kHz rates. The best available sensitivities for interferometers are in the order of 100 MHz, limiting their use to higher velocity measurements. This limitation has apparently discouraged the use of Fabry-Perot systems to any large extent.

To partially overcome the restrictions encountered with the Fabry-Perot interferometer, Smeets and George28-31 developed a more sensitive Michelson (Section 2.4.1) interferometer/spectrometer, that allows the direct measurement of Doppler frequencies to resolutions limited only by the stability of the laser itself. The system is illustrated in Figure 4.7. A single laser beam illuminates a sample volume which contains moving SCs. The radiation scattered by the SCs is Doppler shifted by an amount that is proportional to the velocity component in the k⊥ direction. The scattered radiation is collected by a lens, and is then conducted to the interferometer by a single mode optical fiber.

The light bundle leaving the optical fiber is collimated with an objective lens. After having been linearly polarized, it passes through a Pockels cell and then enters a Michelson interferometer. The Michelson interferometer has a number of important features: 1) its two legs have a path difference; 2) it employs a polarizing beam splitter to direct one polarization into the short leg and the other into the long leg; and 3) it employs a second polarization beam splitter to produce two interferences that are complementary to each other. The relative phase of the two complementary signals can be adjusted so that they cancel. A change in wavelength of the incoming light wave can be exactly compensated for by a phase shift produced by the Pockels cell. In this manner, the velocity is proportional to the voltage required on the Pockels cell to maintain the null condition.

The velocity range of this technique has essentially no upper limit, but, as discussed in Chapter 1 (Section 1.5.2), lasers have inherent frequency instabilities that will limit the lowest velocity measurable. With some degree of stabilization, frequency can be stabilized to one part in 104, representing a low end value of 6 MHz for the 0.5 micron wavelength. This translates into a minimum velocity of 3 meters/second without additional laser stabilization.
Figure 4.6 Laser velocimeter employing Fabry-Perot interferometer.

This LDV represents a significant advance and should open the way to a wide variety of new applications. Unique advantages include the following:

- No upper limit on velocity range
- Fast, direct, and continuous velocity readout
- Works well with heavily seeded flows
- Works on large particles and surfaces
- Much simpler transmitter, less affected by density gradients and lower laser coherence than other methods
- Easily adaptable to fiber optics

4.1.2.2 Heterodyning Methods - Equation 1.7 describes the intensity of the sum of two waves that have different frequencies. Namely, when two such waves are combined, the resulting time averaged intensity, averaged over times that are large compared to the optical period, is modulated by a beat frequency that is the difference between the two original frequencies. If the frequencies are sufficiently close to each other, then the resulting beat frequency can lie in a range that is measurable by conventional electronics. This is called heterodyning and is routinely used in electronic signal
processing. Therefore, to measure Doppler shift electronically, one can mix the laser beam coming from the laser, with the Doppler shifted, scattered radiation to produce a beat signal at the Doppler frequency. Commercial electronic equipment now measures up to about 200 MHz, which covers, essentially, the velocity range of interest to aerodynamicists.

The different methods of heterodyning and the place at which heterodyning actually occurs create distinctively different systems and system requirements. The most important of the distinctions is that between systems which employ coherent detection or incoherent detection. Figures 4.2, 4.5, 4.8, and 4.9 show the differences.

With the systems of Figure 4.8 and 4.9, the Doppler shifted radiation is mixed with unshifted radiation. This has been termed the reference/scatter (RS) mode. The amount of the Doppler shift is given by Equation 4.1 for both cases, but there is an important difference in the actual physics of the two cases. The system of Figure 4.8 has two real wavefronts entering and creating real fringes in the sample volume. Anything that happens to the radiation after leaving the sample volume will happen identically to both beams, since they are overlapped. In this sense, the heterodyning has already taken place at the sample volume and the system can be treated as an incoherent system thereafter.
Contrast this with the reciprocal system of Figure 4.9. No real fringes exist in the sample volume. The alignment and coherence of the two beams is extremely critical all the way to the photodetector where heterodyning occurs.

Figures 4.2 and 4.5 illustrate the types of systems that heterodyne two different Doppler shifted waves to produce a beat frequency. If the two different portions of the wavefront have Doppler shifts equal to

\[ f_{D1} = \frac{V}{2\lambda} \cdot (k_{s1} - k_{s0}) \]  
\[ f_{D2} = \frac{V}{2\lambda} \cdot (k_{s2} - k_{s0}) \]

then the associated intensity variation of the mixed waves is modulated by the difference frequency

\[ f_{D} = \frac{V}{2\lambda} \cdot (k_{s1} - k_{s2}) \]

which is proportional to the velocity component lying along \( k_{s1} - k_{s2} \) direction, which is independent of the illuminating beam direction. This is the so-called dual-scatter (DS) mode, also referred to as the Doppler difference mode. Note that the system of Figure 4.2 is input aligned, while that of Figure 4.5 is output aligned. The DS system, with real fringe or input aligned optics, has become, by far, the most widely used configuration for a variety of reasons to be discussed.

4.1.2.3 Signal Analysis and Interpretation - Since the elementary form of the Doppler signal does not usually exist in practice, we must expand the discussion to cover signal characteristics that are commonly encountered. Most of the complicating effects can be described in terms of the fringe model.

The point SC, passing through the sample volume of a DS system, scatters light in all directions with an intensity form illustrated in Figures 4.4 and 4.10, that depends upon where it traverses the sample volume. A path through the center of the sample volume produces a sinusoidal signal enveloped in a Gaussian. The Fourier transform of the signal contains two major components, first, a lower frequency part of the spectrum that makes up the Gaussian part of the signal, that is commonly called the signal "pedestal" or sometimes (incorrectly) the DC part of the signal and, second, the frequency of the sinusoid which is the Doppler part of the signal. It is important that these two parts of the signal be separable, and this can be designed into a system.

The signal form can be altered by many practical effects presenting the following questions:

1. How is the signal affected by
   (a) finite size SC?
   (b) more than one SC at a time?
   (c) different velocity SC at one time?

2. When does the fringe model break down?

3. What are other broadening effects?

If the SC has a finite size, then the scattered light signal can never be zero, since some part of the particle must be illuminated at all times, even if another part of the particle is in a dark fringe. This reduces signal contrast or signal "visibility". The visibility parameter has actually been used as a particle sizing parameter, but it has not proven reliable and has been abandoned for this purpose by most investigators.

A second effect of finite particle size is the failure of the fringe model to predict the signal character for large angles between the two illuminating beams. Finite sized SCs exhibit a larger forward scattering component, so that the light scattered from two wide-angled beams will differ in every viewing plane except the one that bisects the angle between the two beams. This will lead to a second reduction in signal contrast that is not predicted from the fringe model.
Signal Fourier Transform Sources of Signal

- Point S.C. through sample volume center

1. S.C. size comparable to fringe space
2. Unequal beam intensities
3. Wide angle between beams plus large S.C.

1. Electronic noise
2. Multiple S.C.

Figure 4.10 LDV signal types - dual scatter system.

Inherent in the discussion so far is the requirement that the SCs do actually follow the flow. Large SCs may not be capable of following the flow. The subject of the seeding of the flows has become so dominant in reliable LDV that this will be treated separately.

More than one SC in the sample volume affects the two different LDVs, reference scatter and dual scatter, differently. This is further compounded by the existence of different size SCs, some of which follow the flow and some of which do not. Finally, the actual fluid velocity itself can change in space and time within the sample volume. All of these result in Doppler frequency broadening, and usually in a decrease in the signal-to-noise ratio. Typical signal types for reference scatter systems are illustrated in Figure 4.11.

Isolating broadening effects, to determine which ones are caused by the instrument and which ones are caused by the flow, is one of the major problems of the LDV signal interpreter. For example, optical system broadening occurs when the fringes are not parallel and equally spaced in the sample volume. This can be caused by poor optics or by system misalignment. Electronic broadening is caused by the instrument incapability to respond to the required frequency. These are usually characterized by first using the instrument on a known, fixed velocity.

The signal-to-noise ratio is an important parameter to characterize an LDV. This can be estimated by assuming that the major noise source is the photodetector dark current noise. In general, other sources of noise, such as laser noise and ambient light and electronic noise, must be considered. The average intensity of light striking the SCS is given by

\[ I_0 = \frac{8P}{\pi D_g^2} \]  

(4.7)

where

- \( P \) is the laser power
- \( D_g \) is the diameter of the laser beam at focus

The amount of scattered radiation is equal to

\[ I_S = \sigma(s)I_0\omega_{SC}/R^2 \]  

(4.8)

where
Signal Fourier Transform Source

1. Single point SC through sample volume center

1. Multiple SC in sample volume

1. Electronic noise

Figure 4.11 LDV signal types - reference scatter systems

- \( \sigma(s) \) is the scattering cross section
- \( N_{SC} \) is the number of SCs
- \( R \) is the distance to the collector.

Brayton\textsuperscript{32} has calculated the S/N for the two types of LDVs as the following:

\[
(S/N)_{RS} = \frac{\sigma^2 n_0(s) P N_{SC}[\lambda \theta]^4}{4hcB} \tag{4.9}
\]

\[
(S/N)_{DS} = \frac{\sigma^2 n_0(s) P[\lambda \theta]^4}{16hcB} \tag{4.10}
\]

where

- \( \theta \) is the scattered light collection angle
- \( n \) is the detector quantum efficiency
- \( B \) is the system bandwidth
- \( h \) is Planck's constant
- \( c \) is the light speed
- \( \lambda \) is the wavelength

Especially noteworthy is the S/N improvement with collecting aperture which favors the DS system. Note also that the RS system S/N improves with the number of SCs. Finally, it must be noted the S/N decreases with velocity because the required system bandwidth, \( B \), increases with velocity.

By their very nature, LDVs produce discontinuous, digital-like data because one velocity measurement is produced per particle transit at the very most. These data points must be properly treated statistically to arrive at meaningful interpretations. For example, in a highly turbulent flow, more high-speed data samples will be collected simply because they arrive at the sample volume more often, so one might statistically bias the data to an incorrectly high mean velocity. This bias error is easy to correct and the time averaged velocity can be calculated by...
where \( \langle V \rangle \) is the velocity of the \( i \)th particle, \( t_i \) is the transit time of the \( i \)th particle, and \( T \) is the average transit time. On the other hand, the reduced S/N for high-speed SCs can bias the data towards low velocity.

If a particle size distribution exists and if all of the particles do not track the flow, the problem of interpretation of turbulence data is virtually impossible.

4.1.3 Optical Geometries

4.1.3.1 Systems - The optical requirement of a DS system is that a sample volume of desired dimensions be produced by focusing and crossing two laser beams at a desired point in space, followed by collection of the maximum amount of scattered light from the volume and from nowhere else. This can be reduced to the following optical functions (refer to Figure 4.12 and 4.3).

![Figure 4.12 Laser velocimeter geometry.](image)

1. Referring to Chapter 1, determine the F/number required to produce the (transmitter defined) sample volume dimension and location, then expand the laser beam to the required input diameter. This is called beam conditioning.

2. Produce two parallel, collimated beams of light, separated by the desired amount (see Figure 4.12).

3. Focus and cross the two beams at their focus, by passing through a high quality lens or mirror or by some other means.

4. Collect the light scattered from the sample volume with an efficient receiving system that is stopped to further define the geometrical sample volume.

A wide variety of beam splitter elements have been used to produce the two parallel beams of light. Figure 4.13 illustrates some of the most commonly used elements (see also Figure 4.20). With elements like 4.13a-b, the angle between the two output beams can be adjusted by rotating one of the components. Although sometimes useful, this can add to the system alignment problem. The remaining beam splitters automatically transform a single input beam into two parallel beams (if the components are accurately made). Note that, of the remaining components, only c, r, and h have equal pathlengths for the two beams. This can be a critical point for lasers with a short coherence length, and pathlength compensation may be provided by additional components. For example, (b), (d) and (e) can be used in combination to produce two collimated beams of equal pathlength, as is shown in Figure 4.14.

Other types of beam splitters produce two waves that diverge from a point, and these waves can be reconverged to produce a detection volume that contains interference fringes. The two beams can be produced by an amplitude or phase grating, or by a holographic optical element (see Figure 4.15). This can be understood simply as the imaging of the grating into a detection volume. If the grating moves, then the fringes likewise move, providing a useful mechanism for determining the direction of the velocity of the scattering center. An SC that moves against the direction of the fringe motion is associated with a higher frequency than one that moves in the direction of the fringe motion.

Acousto-optical cells can be used to split the beam into two components, each having slightly different frequency. When the two beams are reconverged, the fringes will move as described in Chapter 1. This can also be understood as the imaging of an initially moving set of fringes.
Another beam splitter in this class is the Wollaston prism, which produces two beams that are polarized normal to each other; so, when the two beams are reconverged, interference does not occur. The procedure is illustrated in Figure 4.16. Polarizers can be placed with the polarization direction at 45 degrees with respect to the two crossing beams so that interference can occur and virtual fringes are produced. If two different polarizers are used with a 90 degree difference in their respective polarization directions, then the two resulting beams will produce signals that are 180 degrees out of phase with each other. These two can be mixed electronically to remove the pedestal from the signal.
4.1.3.2 Two Component Dual Scatter Systems - Figure 4.17 illustrates the basic two component DS system. One can interpret the process as the production of two sets of fringes in the detection volume. The DS system can take on a second form illustrated in Figure 4.18. In Figure 4.18, the scattered light is divided with a beam stop into two sets of beams, each having an associated set of virtual fringes. In Figure 4.17, the fringes are real.

The only added problem, in going from one to two components of velocity, is the separation of the two components upon detection. The separation can be done by polarization, by color, by electronic means, or even by coherence discrimination, but in recent years most investigators have applied color separation. Argon lasers are ideally suited for this process, since two dominant wavelengths (488 nm and 514 nm) are of nearly equal power in available lasers. In Figure 4.17, beams 1 and 2 are of two different colors, while beam 0 is a mixture of the two.

In a two component system, the orientation of the fringes, with respect to each other and with respect to the flow direction, can be optimized for most cases. If the flow direction is approximately known, then the fringes should be oriented symmetrically about this direction. Neither fringe set should be placed with fringes parallel to the velocity direction, for then the SC will not cross any of the fringes. The optimum angle between the fringes is rarely 90 degrees, and is more likely to be about 60 degrees to get the most accurate data for both components of velocity.

4.1.3.3 Three Component Dual Scatter Systems - Three component systems present a number of unique and difficult conditions for the designer. The problem is to produce three independent sets of fringes in a sample volume, and then to separate the scattered light signal into components that can be associated with the correct velocity component. Practical difficulties arise from the lack of a good laser candidate that has three different wavelengths of equal intensity, and from the need to orient the fringes so that the SC crosses a sufficient number of fringes in each of the three sets of fringes.

The problem of producing three fringe sets of equal intensity has generally been solved, either by using two lasers (for example, an argon-ion laser and a krypton laser) to produce three wavelengths\(^1\), or, by tuning down the power in the two principal wavelengths of argon lasers and choosing a third, say at 470 nm (see Figure 4.19). Either method leads to a rather difficult and expensive configuration. Pfeifer\(^2\) (1985) has presented a novel approach to this problem, in which the argon laser is used partly to pump a dye laser for one of the components, while using the two principal wavelengths for the other two components.

---

\(^1\) Three component systems present a number of unique and difficult conditions for the designer. The problem is to produce three independent sets of fringes in a sample volume, and then to separate the scattered light signal into components that can be associated with the correct velocity component. Practical difficulties arise from the lack of a good laser candidate that has three different wavelengths of equal intensity, and from the need to orient the fringes so that the SC crosses a sufficient number of fringes in each of the three sets of fringes.

The problem of producing three fringe sets of equal intensity has generally been solved, either by using two lasers (for example, an argon-ion laser and a krypton laser) to produce three wavelengths\(^1\), or, by tuning down the power in the two principal wavelengths of argon lasers and choosing a third, say at 470 nm (see Figure 4.19). Either method leads to a rather difficult and expensive configuration. Pfeifer\(^2\) (1985) has presented a novel approach to this problem, in which the argon laser is used partly to pump a dye laser for one of the components, while using the two principal wavelengths for the other two components.
4.1.3.4 Moving Fringes - In any of the geometries described above, a bias frequency can be added to the Doppler signal. In terms of the fringe model, this can be described as producing a system with moving fringes, so that for a stationary SC in the sample volume, the scattered light will be modulated at a frequency determined by the rate at which the fringes move over the SC. Fringes that are produced by interfering beams of two different optical frequencies will move at a rate that increases with the frequency difference. Any point in the fringe pattern will vary sinusoidally at the frequency difference of the two waves. Moving fringes can be produced by several different methods. A frequency shift can be added to either or both of the two waves with a Bragg cell. If the fringes are produced by using a grating, then the fringes can be moved by moving the grating. In practice, this is implemented by using a radial grating on a rotating disk.

4.1.3.5 Fiber Optics - Fiber optics can be used to conduct the light from the laser to the transmitter. If the transmitting fiber is a single mode fiber, then the end of the fiber is treated as a point source, just as it would be if it had been produced by focussing the direct beam from the laser. It is then collimated and split in any of the ways described above. Alternately, light being conducted in a fiber can be split, as shown, by a fiber coupler into two beams that can subsequently be collimated into parallel beams. (Figure 4.20)

Single mode fibers can conduct no more than about 100 milliwatts of power. Above this power, nonlinear optical interaction with the material can cause prohibitive losses. If greater power lasers are used, then the light must be transmitted through coherent fiber bundles. The output beam is no longer a simple diverging wavefront, but contains speckles. This somewhat complicates the sample volume, first by enlarging it and by producing patches of fringes that are not in phase with each other.

The scattered light can be collected by a lens that focusses the light into a fiber conducting it to a remote location into a photodetector. When incoherent detection or heterodyning at the SC is involved, this fiber can be a much larger fiber since coherence is no longer required (Figure 4.21).
4.1.3.6 Reference Scatter and Other Direct Doppler Systems - Direct Doppler systems rely on the measurement of the Doppler shift of the scattered light (as opposed to the difference in Doppler shift used in dual scatter systems). Only one beam need be projected into the flow, and the geometrical sample volume, which lies along this transmitted beam, is defined by the overlap of the image of the receiving aperture projected across the transmitted beam (illustrated in Figure 4.22). The properties of such a system can actually be approximated by using the fringe model of 4.1.1.2, wherein the virtual fringes are produced by an imaginary beam of light, originating at the receiver aperture and projecting back through the system optics to cross the transmitted beam from the laser (i.e., the reciprocal system).

To determine the amount of the Doppler shift of the scattered light, either an optical spectrometer/interferometer can be used directly or a heterodyning technique can be used to beat down the frequency of the signal to a frequency that is measurable electronically.

The heterodyning procedure requires a reference wave be mixed with the scattered light. An optical system that is similar to the DS systems already described can be used to achieve this as shown in Figure 4.12. The beam splitter is designed to produce a reference wave that is typically 3 to 30 times as bright as the scattered wave, and, for this configuration to perform well, the reference wave must pass through the flow field largely unscattered and unmodulated. This arrangement is particularly convenient in terms of alignment, since the reference wave, which passes through the sample volume, is automatically aligned with the scattered wave. This is sometimes referred to as a self-aligned or input aligned system.

In highly seeded or highly turbulent flows, the configuration of Figure 4.12 may not be applicable, since the reference wave can contain a large quantity of scattered light and phase noise that can reduce the signal-to-noise ratio. Therefore, in some cases the signal can be optimized by passing the reference wave around the flow field. Precise alignment of the reference wave with the scattered wave in such a system, is extremely difficult. As described in Section 4.1.1, it is important that the reference wave and scattered wave overlap precisely over the collecting aperture. This requires the F-numbers of the two lenses to be the same. This is a simple design problem for the system of Figure 4.12, but is more difficult when the reference and scattered waves emerge from two different points.
4.1.4 Electronic Signal Processors

4.1.4.1 Oscilloscopes - The simplest form of signal processor is an oscilloscope that is set to trigger at some threshold. If the signals generated by different SC are similar, then a sine wave with a Gaussian envelope will be observed, and the frequency can be determined simply by measuring the scope screen. If the signals from different SC are not similar, then a storage scope must be used. The scope is set to trigger on the smallest possible burst signal, allowing the examination of individual "Donpler" bursts. The bursts will have the appearance shown in Figure 4.10.

Using an oscilloscope is a good way to evaluate the quality of the overall system, but it is not a practical method to collect a useful amount of data. The signals should exhibit high contrast and low noise. The measured amplitude will help in determining the threshold of other types of electronic processors. Some processors actually use the scope trigger to trigger the processor that analyses the Doppler bursts. Low contrast signals are a symptom of poor system alignment, incorrect beam intensity ratios, too large SC, or insufficient electronic frequency response, or combinations of these.

4.1.4.2 Spectrum Analyzers - The first signal processors to be used in LDV were spectrum analyzers. Even though they are simple to use, their limitations severely restrict their usefulness. The signal from the photodetector is fed directly to the analyzer, which sweeps a variable electronic filter over a preset frequency range, and presents an amplitude versus frequency plot on a cathode ray tube at rates ranging from seconds down to milliseconds. When used in LDV, the sweep rate must typically be quite slow to allow enough Doppler bursts to enter the system at each frequency. The system output is broadened electronically by the signal envelope, by noise, and by phase fluctuations caused by more than one SC in the sample volume at one time. Therefore, it is difficult to apply the instrument to anything other than laminar, constant velocity flows. Spectrum analyzers are useful to determine the frequency range in which other electronics must operate.

4.1.4.3 Frequency Trackers - If the Doppler signal were merely a frequency modulated sinusoid with constant amplitude, conventional electronic frequency to voltage conversion would be an acceptable method of signal processing. However, in special instances, a conventional frequency tracker is incapable of dealing with large amplitude fluctuations, noise, and especially signal dropout. Therefore, the signal must usually be conditioned before it is analyzed, and signal dropout must be considered.

Frequency trackers inherently require large amounts of seeding to provide a nearly continuous signal. They are most useful in the LDV of liquid flows where it is easy to provide the required seeding. They have virtually vanished from use in gaseous flows. LDV in gaseous flows is basically a binary process, while frequency tracking in the usual sense is an analogue process. Because of the limited use, they will not be discussed further and the reader is referred to the references for further information.

4.1.4.4 Electronic Counters - Counters have evolved to a relatively uniform set of electronic functions. These are illustrated in Figure 4.23. The signal entering the processor is a sinusoid with a Gaussian envelope plus a vast range of noise. The signal is first high pass filtered to strip away the pedestal, leaving a signal that oscillates about zero. Next, the signal is bandpass filtered to remove noise that lies outside the frequency range of interest. The signal-to-noise ratio is inversely proportional to the bandwidth, so it pays to select this as closely as possible. Validity checks are done at this point to insure that the signal has the correct appearance. For example, a condition for the signal to be accepted for processing is that the maximum or minimum is observed, the signal must be zero before another maximum or minimum is observed. This is called a "three-level check". The signal is further amplified and clipped so that it becomes a binary signal.

In some systems, additional checks are done at this point to insure that the frequency of the signal does not change during a Doppler burst. This is done by counting the time required for a set number of cycles occurring early in the burst, and comparing it with the time required for the entire burst. When divided by the total number of cycles, this number should be the same to within a small percent. Otherwise, the signal should be rejected as invalid.

4.1.4.5 Simultaneity Test - Multiple component measurements often require that valid signals be observed simultaneously from the different components for the data to be useful. Therefore, coincidence circuits reject signals that do not occur simultaneously in all of the components. Almost without fail, this severely reduces the data rate. Fairly simple electronic circuitry is used for this function. When a threshold is exceeded in either channel, a pulse is sent to an AND gate in the other channel. The AND gate remains both the pulse unless both input gates are open. A preset time window can be produced in this way allowing the acceptance of the two velocities. By increasing the pulse duration, the data rate can be increased at the expense of precision. The optimum pulse duration is a function of the turbulence frequency in the flow and, to be meaningful, the measurements must be made to within times that are small compared to the characteristic times for the flow.
4.1.4.6 Transient Recorders - Today's digital transient recorders can digitize an analog signal at rates in excess of 200 MHz, so it is possible to store the entire Doppler burst in digital form. Then the stored signal can be processed by an ordinary computer. The memory of such a system will be filled quickly, limiting the use of the recorder to short times.

A variety of methods have been devised to extend the use of such systems to longer times. Data compression schemes can insure that the recorder accepts data only when a Doppler burst is present. Buffer memories are also used to dump the recorded data at slower rates as the recorder memory is filled. This provides a sampling of the flow velocity with dead times lasting a few milliseconds for each data dump.

Use of a transient recorder is especially important for two specific cases, one when the signal is extremely poor and requires additional processing to extract usable signals, and a second when the quantity of data is so small that every signal is needed to constitute enough data. Having the entire signal for processing at a slower rate permits the use of powerful filtering procedures such as correlation. Portions of the signal can be correlated with itself to improve the signal-to-noise ratio. Another procedure is to Fourier transform the signal to determine its strongest spectral content. Once the information signal is identified bandpass filtering can be applied around the frequency of the desired signal to further remove the noise in the superfluous frequency band.

4.1.4.7 Correlators - The previous section described how correlation is used to improve the signal-to-noise ratio of a stored signal. High-speed digital correlators have been developed to perform this function on the live Doppler bursts. The fastest such machines can correlate signals of frequency in excess of 100 MHz. If the Doppler bursts are all of the same frequency as in a laminar flow, then nearly perfect Doppler bursts can be constructed from an extremely noisy signal, and the correlation time will last as long as the Doppler burst. If the frequency of the Doppler bursts varies as in a turbulent flow, then the correlation time will provide a measure of the turbulence frequency of the flow. Correlators are commonly used in front of burst counter processors.

4.1.5 Applications

Hundreds of interesting and impressive applications of LDV could be cited; however, the objective here is to illustrate the principles, the state of the art, and to describe some recent developments that have not yet reached textbooks on the subject. Therefore, the approach taken is to pick a few examples from the various laboratories visited during preparation for this book that illustrate the methods and system variations especially well. Examples are given for use in wind tunnels, shock tubes, turbomachinery, and in the atmosphere.
4.1.5.1 Wind Tunnels - ONERA (France): Excellent examples of elaborate LDV systems are those that have been developed by ONERA in France\textsuperscript{42-43} for use in the Modane and Le Fauga wind tunnel facilities (Boutier et al. 1983). The S3 Mondane facility is a blow-down transonic tunnel with a run duration of less than 200 seconds. The velocimeter is a two component forward scatter system using an Argon laser at 9 watts, blue and green fringes with spacing of 24.5 micrometers, oriented at 45 degrees with respect to the tunnel axis, and modulated with Bragg cells to 5 and 6.25 MHz respectively. The forward scattered light is collected by a Cassegrainian telescope that has a 1.2 meter working distance at f/6.

Figure 4.24 is a diagram of the systems, showing the extensive use of computers to command and track the independent traversing of the transmitter and receiver units in three dimensions. This system can take measurements at one point and then move to a different point, for a second measurement, in a few seconds. Incense seeding is done to increase data rate and to insure that the particles follow the flow. The system has been used successfully to study boundary layers, wakes, and flow profiles.

![Figure 4.24 General diagram of a two-component LDV system.](image)

In continuous tunnels, like the S3 Modane and F1 Le Fauga closed circuit facilities, where more time is available for collecting data, backscatter systems can be used. Such a system is shown in Figure 4.25. At higher velocities, even here, forward scatter systems have been required. In pressurized systems such as the Le Fauga facility, where it is required to place the transmitter in the pressurized plenum, another problem occurs. At higher pressures, the Brewster angle in the laser changes due to the increase in the refractive index inside the laser cavity. This problem has been solved here by placing the laser in a sealed box.

The most recently developed system of ONERA is the three-component system diagramed in Figure 4.19. The system uses three beam pairs at different wavelengths. These are extracted from two different argon lasers. Again, the transmitter and receiver are traversed under computer control. This is one of only a very few three-component systems in operation.

LDV systems that are similar to the ones in use at ONERA are in use at other research centers. A number of these are described in more detail in NASA Conference Publication 2243 (1982).\textsuperscript{4} Some examples are noted in the following:

NASA Ames Research Center (USA): Two-component LDV systems are used in various tunnels such as the 13 by 26 meter subsonic, the 2.5 by 3.3 meter transonic, and the 2 meter supersonic tunnels. Backscatter systems are primarily used. A long-range LDV
system has been developed for the 13 by 40 meter subsonic wind tunnel, which is the largest such test facility in the free world.

**NASA Langley Research Center (USA):** Two-component backscatter systems for the 5-meter transonic tunnel use a two-dimensional traverse and a zooming system for the third dimension. A two-component backscatter system is used in the 4 by 7 meter subsonic tunnel. This is a somewhat unique system in several respects. It employs what is probably the largest two-dimensional traverse for LDV, covering a sampling volume of 1 by 2 meters and zooming for the third dimension. The traverse alone weighs in excess of 2700 kg. A tilt and pan system also allows rapid scanning of the transmitted beam in addition to the traversing action. A 12-watt argon laser is used. These two systems are permanently dedicated installations. Other smaller systems are employed in tunnels ranging up to hypersonic Mach numbers.

**DFVLR (Germany):** A two-component backscatter system is used in the 3 by 3 meter subsonic tunnel and in smaller facilities.
NASA Lewis Research Center (USA): A two-component backscatter LDV has been used in the 2.7 by 3 meter transonic tunnel. This system uses a two-dimensional traversing system with zooming optics and a 15-watt argon laser. The system is fully enclosed in a water-cooled, vibration isolation housing.

Naval Surface Weapons Center (USA): A two-component backscatter LDV is used in low velocity flows in a 1 by 1.3 meter tunnel; a two-component forward-scatter LDV has been used up to Mach 5; and a three-component system has been used at Mach 3.375.

U.S. Air Force, AFEDC (USA): Two and three-component LDV systems have been used in a variety of tunnels ranging in test section diameter up to 5 meters and in Mach numbers up to Mach 3.5. Systems at ARM hangar are probably the highest airspeed systems currently in use. This processor, which has been called a "real time" Fourier transform signal processor digitally records the entire Doppler signal, analyses the information at rates in excess of 20 MHz. This is essentially a specialized digital recording system that has eliminated the problem of limited memory described above.

U.S. Air Force, WPAFB (USA): A two-component backscatter LDV system has been used in a number of wind tunnels up to 1 meter test sections at Mach numbers up to 6.5-8.

ISL (Germany/France): Two-component forward and backscatter LDV systems have been developed for use in a variety of tunnels. A different approach to the problem of obtaining three wavelengths for three-component systems has been developed by Pfeiffer (1985). In this system, a 10 watt argon laser is used to provide two of the wavelengths, and also to pump a 660 nm line from an 860 dye laser to provide CO2. Also, a unique traversing corner reflector system makes traversing faster and easier than traversing the whole LDV system. Most large LDV systems require some adjustment in the output beams to insure that they cross and focus at the same point in space. A recently deployed ISL system employs a segmented output lens, with adjustable segments that provide a convenient method for making this adjustment.

4.1.5.2 Shock Tubes - Shock tubes present a difficult applications environment because of the high velocities and short run times. Such facilities are usually high pressure facilities requiring remote operation of the system.

A system for shock tubes and outdoor shock phenomena, developed by Spectron Development Laboratories, has been used in the 3 meter and 7 meter CERF shock tubes at Kirtland Air Force Base and on the White Sands Test Range. This forward scatter system employs fiber-optics to conduct the light from a 4-watt argon laser to four sample volumes located between two blast wings. The scattered light from the virtual fringes are produced by the fiber-optics to the photomultiplier tubes, which are located up to 300 meters away from the sample volume. In such transient applications, one cannot afford to reject any of the signal; the entire signal is recorded on three recorders.

Sweeze13-10 has applied an LDV using direct Doppler frequency measurement with the Michelson spectrometer (described in Section 4.1.2.1) in a variety of flow situations, including shock tubes. Figure 4.26a and b illustrates the setup. The transmitting and receiving optics is extremely simple for this type of system, augmenting its use in severe environments. The velocity components under observation can be discerned by examining the directions of the virtual fringes. The fringes are produced by the transmitted laser beam and a second beam that is projected toward the receiver. So, for the case illustrated, the fringes for the s velocity component are perpendicular to the flow axis.

The innate high-frequency response is illustrated in the data samples shown which also illustrates that the system is most readily used to measure velocity change, making it an excellent candidate for turbulence measurement.

4.1.5.3 Long Range Systems - LDV systems have been used at ranges up to several hundreds of meters in wind sensing. Some of the earliest work in this application employed argon lasers at 488 and 5145 nm and CO2 lasers.43 The basic configuration for a long-range fringe system is illustrated in Figure 4.27. A Cassegrainian telescope is used to project and cross two laser beams. The backscattered radiation is collected either by the same telescope or by a second receiver system that sits off-axis from the transmitter. Earlier systems employing standard signal processing methods required large SCs to work at long range. More recently, the successful application of correlators has significantly improved the performance4.67 Durst et al.41 use a rotating grating beam splitter, a separate transmitter and receiver, a 7-watt argon laser, and a correlator in the signal processor, have established that such a system is applicable to a range of about 500 meters.

Atmospheric perturbation of the beams is quite serious at visible wavelengths at long ranges. This provides incentive to go to longer wavelengths. Mid-infrared lasers are available at much higher powers. Perhaps of even greater practical importance is that the 10.6 micrometer wavelength of CO2 is eyesafe at much higher powers. Such systems are operated in the direct Doppler mode and measure the velocity component along the projected beam. The systems, having been operated at ranges up to several kilometers, hold great promise for use in aircraft for the detection of windshear and clear air turbulence.
4.1.5.4 Fiber-Optical Systems - LDV systems employing fiber-optics are rapidly gaining in use. An optical transmitter and receiver can be located near a measurement volume, while the laser and electronics are located many meters away, without much penalty in laser power. Equally important is the use of fiber-optics to miniaturize part of the LDV system.

Two systems that illustrate these features are presented. The first is the so-called LDV endoscope of Reithmueller, illustrated in Figure 4.28. The light from a 2-watt argon laser is focused onto the end of a 100 micron diameter multimode fiber that conducts it to the measurement head. At the measurement head, the light diverges to a collimating lens that produces a 4 mm diameter beam. This beam is then split into two parallel beams in the usual way (Figure 4.13). A focusing lens then focuses and crosses the two beams. The backscattered light is collected by the same lens, which then focuses it onto a receiving fiber. The fiber then conducts it to a photodetector.

In this system design, fiber-optics is exploited principally to produce a small, lightweight measuring head that can be traversed independent of the laser and the photodetectors. These elements, which are usually not only large and bulky but also subject to misalignment and electrical noise sensitivity, can then be located in a well-controlled room that is located in a convenient place. The Von Karmen Institute has at least four of these systems in routine operation in a variety of wind tunnels and other test facilities.
Durst has even further exploited the properties of fiber-optics that allow the miniaturization of a laser transmitting and receiving optical head. Figure 4.29 illustrates the design concepts employed. Figure 4.29a is a one-dimensional system, in which the laser beam is first split into two parallel beams that are frequency shifted by Bragg cells, then focused into two separate fibers. The beams are then transmitted to a miniaturized element that recollimates the two beams before finally focusing and crossing them at the sample volume. The backscattered light is collected by the same collimator, which focuses the light onto a return fiber that ultimately carries it to the photodetector.

Figure 4.29b illustrates how the same concepts can be used to produce a two component miniaturized system. This system uses an argon laser to provide the two colors for the two components. These are collected and carried to the photodetector by a single fiber. At the photodetector, the light is split into the two components by a dichroic splitter. These systems have been constructed in almost pencil-sized optical heads, small enough to be inserted into the head of an internal combustion engine.

4.5.5.5 Turbomachinery and Internal Combustion Engines - The ability to produce rugged LDV systems has expanded their use in the harsh environments presented by internal combustion applications. One example is provided by the fiber-optic LDV system of Durst described in the last section. A major problem for this application has always been the problem of providing optical access to the inside of the engines. This LDV is sufficiently small that it could be inserted into a small spark plug port. Swirl velocities were measured at several points in the cylinder, ranging up to about 10 meters per second.
Optical access methods also include using glass or quartz cylinders and installing windows into the cylinders at various points. Extensive study of the windowing techniques for such engines has been made at the U.S. National Combustion Laboratory at the Livermore Laboratories. Schock has developed a unique approach to the problem of producing a good sample volume through the otherwise aberrating cylindrical walls of a quartz cylinder. Using a holographic optical element designed to focus and cross two laser beams through the cylindrical walls without aberration, he was able to remove the astigmatism in the beams, producing a diffraction limited focus. The principle is illustrated in Figure 4.30. The HOE is produced by forming a hologram of a pair of laser beams that have diverged from the point through the cylindrical walls. When the hologram is illuminated again, the two beams have the necessary negative astigmatism to exactly correct that introduced by the walls as the two beams pass through.

In an extremely efficient system, Pfeifer and Schafer have used the concept illustrated in Figure 4.31. A spherical mirror was used to collect the transmitted, unscattered light from the sampled volume, and then to reimage it back into the flow at a slightly displaced location. In this way, two sample volumes with adjustable spacing were produced with only one transmitter. The two receivers are then placed on either side of the flow.

A more elaborate, two-component, two-point LDV produced at ONERA is illustrated in Figure 4.32. This system employs two colors to produce the two components, and two independent sample volumes are positioned by computer controlled traverses.

4.1.5.7 Flow Seeding - The subject of proper seeding of the flow has been acknowledged as one of the critical problem areas in LDV. Fundamental to all applications of LDV is the assumption that the flow can be characterized by the motions of finite sized particles, that either reside in the flow naturally or are introduced into it artificially. Apparently, much of the published data taken with LDV is in error because of the breakdown of this assumption. The problem is considered of sufficient importance that entire conferences have been devoted to the subject (see, e.g., NASA Conference Publication 2393). Nicholas has summarized the effects of particle dynamic effects on LDV measurements in the following way:

1. Interpretation of LDV measurements requires the capability of analytically examining particle dynamic effects.
2. Broad size distributions produce artificial turbulence and should be avoided.
Figure 4.28 LDV Endoscope of Reithmueller. The installation in a cooling tower test shows how the laser and photomultiplier can be located in a control room with the light being transmitted to the optics head through fiber-optics. The optics head can be easily positioned now because of its small size and weight.

(3) Maximum acceptable SC size for typical AEDC LDV mean velocity measurements is less than 0.5 microns.

(4) Absolute maximum acceptable SC size is probably driven by turbulence, not mean-flow requirements.

(5) Correction for mean-flow SC lag may be feasible for monodisperse seed particles.

Figure 4.33 presented in Nichol's paper illustrates the response of various size particles to a shock wave, showing that even the smallest particles required a considerable relaxation distance to attain the gas velocity after passing through the shock.

Different approaches to understanding the seeding problem include drag computation, measurement of monodisperse particle velocity response to various known flow conditions, and response of particles to shock waves.

The approaches to dealing with the problem experimentally include seeding with known particle distributions or preferably with monodisperse particles, and also attempting to reject data arising from particles that are too large to accurately follow the flow. In the early days of LDV, attempts were focused on making measurements without artificial seeding. Today, it is almost universally accepted that starting with a clean flow and artificially seeding with a proper material is preferred.
Although the detailed discussion of seeders is beyond the scope of this treatment, some of these will be mentioned. The reader is referred to the above mentioned NASA publication for a detailed discussion of seeders. Crossway has provided a discussion and characterization of various seeders. Figure 4.33 summarizes some of the seeding methods. One of the most effective is the so-called vaporization condensation seeder using dioctyl phthalate (DOP). This produces nearly monodisperse spheres of submicron diameters. Included in the figure are size distributions of other materials.

Another popular seeder uses Dow Corning polystyrene spheres, which can be obtained in monodisperse solutions with sizes as small as 0.3 micrometers in diameter. The solution is nebulized into droplets that mostly contain only one particle, and then passed through a dryer that removes the liquid by evaporation before passing the particles into the flow. This is practical where small amounts of seeding suffice.
LASER BEAMS FOCUSED INSIDE A CYLINDER

CONVENTIONAL LENS

FOCAL POINT OF GREEN BEAMS IN VERTICAL PLANE

FOCAL POINT OF BLUE BEAMS

MULTIFACETED HOLOGRAPHIC OPTICAL ELEMENT

COMMON FOCAL POINT OF BLUE AND GREEN BEAMS

Figure 4.30 The use of holographic optical elements in LDV to correct aberrations.

PMT (A)

spherical mirror

nozzle

probe volume (A)

probe volume (B)

PMT (B)

Figure 4.31 Optical arrangement for cross-correlation measurements.
4.2 Transit (Two-Spot) Anemometry (Velocimetry)

4.2.1 Theoretical Basis

4.2.1.1 Concept - In simplest terms, a laser beam is split into two components that are focused to two separate points in space, and scatter centers are timed as they transit the two points (Figure 4.35). Even though this is conceptually simpler than LDV, LTA systems require a more complicated optical design to glean all of the advantages of the technique. Injecting all of the laser energy into two extremely small spots exposes scatter centers to the greatest possible intensity, enabling the detection of smaller scatter centers and producing the smallest possible sample volume. A smaller sample volume leads the way to better background noise rejection, and enables measurement closer to light scattering surfaces. As will become evident, the technique is more accurate for both speed and angle measurement. Even so, the method suffers other limitations and it cannot always replace LDV.

Judicious choices of detection and signal-processing methods make possible further enhancement of the signal and reduction of noise. The method is directly adaptable to the use of photon limited detection and represents a limit in small SC detection. So, ultimately, much more is involved than simply timing the transit of SC across the two points. Recently, the concept has been implemented with laser diodes.

The concept of velocity measurement, with a single-point system to produce even greater intensity and smaller sample volume, has been explored without much success. Some of the signal enhancement methods that make two-spot systems work so well cannot be employed with a single spot.

4.2.1.2 Sample Volume - The geometrical sample volume is determined by both the transmitter and the receiver. The transmitter produces two focal regions of Gaussian intensity profile normal to the beam propagation direction and approximately cylindrical in shape. The length of the sample volume in the propagation direction is determined by the receiver. This is illustrated in Figure 4.36. The receiver image overlaps the transmitted focal volume in a way that is determined by the stops used in the receiver. This will be discussed in greater detail below.

4.2.1.3 Nature of the Signal - The greatest S/N can be attained if the intensities scattered from the two spots are monitored separately, and this is the case to be discussed here. With ample illumination, the scattered light signal is Gaussian so that its peak denotes when the SC passes the center of the sample volume inside the flow. There are three possible events: 1) the SC is detected as it passes through the first spot only, 2) the SC is detected as it passes through the second spot only, and 3) the SC is detected as it passes through both spots.

If the signals are observed with an oscilloscope, with the first detector triggering the scope for each detection of an SC and the second detector voltage acting as
PARTICLE RESPONSE TO SHOCK

VEL TUNNEL A $M_{shock} = 4.0$

Figure 4.33a Seeding particles - use in LV.

Figure 4.33b Vaporization/Condensation
Figure 4.33c Collision nebulizer

Figure 4.33d Laskin Nozzle

Figure 4.33e Fluidized bed seeder.
Figure 4.33f Alumina Polishing Powder
Nominal size = 1.0 micron

Figure 4.33h Alumina Polishing Powder
Nominal size = 0.05 micron

Figure 4.33i Cigarette Smoke.

Figure 4.33j Incense smoke.
Figure 4.33k DOP in the Laskin nozzle.  

Figure 4.33l DOP in the Collison nebulizer.  

Figure 4.33m DOP without dilution flow in the vaporization/condensation seeder.  

Figure 4.33n DWP with dilution nitrogen in the vaporization/condensation feeder.  

Figure 33o Ethanol and DOP in the ratio 99/1 in the vaporization seeder.
Enlargement of sample volume showing fringes. Note that the two components are deliberately not orthogonal. The horizontally elongated beams form the vertically elongated part of the sample volume which is used for sizing.

A, B - Green circular beams  
C, D - Blue elongated beams

Figure 4.34 Particle sizing LDV after Hess.

Figure 4.35 Backscatter laser transit anemometer.
the vertical input to the scope, a grouping of signals will be observed at the time $S/V$, where $S$ is the spot separation and $V$ is the mean flow velocity. The group will be narrow for laminar and broad for turbulent flow. The signal will peak when the two spots are lined up with the flow. Other signals resulting from case 1) or 2) above are spread randomly in time and, therefore, do not add up like those for case 3).

A photon limited signal is made of random spikes that do not exhibit a Gaussian profile. The probability is greatest that photons are collected when the SC passes the peak intensity, but they can occur anywhere in the sample volume.

4.2.1.4 Enhancing the Signal with Filters and Correlators - Since the scattered intensity is Gaussian, it is optimum to electronically filter it with a Gaussian filter. The filter must be chosen with a passband that is compatible with the fastest signal.

If the signals from the two detectors are correlated, the randomly timed signals representing cases 1) and 2) above will exhibit little or no correlation, while those from 3) will correlate strongly at time $S/V$ and will be greatly enhanced. Consequently, a signal that may not be observable on an oscilloscope may be observable after correlation.

4.2.1.5 Directional Sensing and Turbulence Measurement - The flow direction can be determined by rotating the two spots until the correlated signal shows a peak at the time $S/V$. If the flow is laminar, this can be determined precisely. If the flow direction is not constant, the peak will appear much broader. If the flow direction is varying rapidly over wide angles, then the number of occurrences of cases 1) and 2) above increases and the number of occurrences of case 3) decreases, resulting in a poor signal. Therefore, the method is not easily usable for highly turbulent flow.

4.2.2 Optical System

A typical optical system is illustrated in Figure 4.37. In the following sections, we shall discuss the selection of optical components and the various options that are possible. The design objective is to concentrate as much of the laser light as possible into two smallest possible spots, to collect scattered light from these two spots independently, to reject light from everywhere else, and to provide for rotation of the spots.

4.2.2.1 Beam Splitting - Several types of amplitude division beam splitters are in use, the two most suitable for LDA being the Rochon and Wollaston polarising beam splitters. Either component produces two collimated beams of crossed polarisation emerging with an angle between them. Other beam splitters are in use and offer advantages of adjustability, but this can cause mechanical inconvenience or unreliability.

The collimated beam from the laser is divided into two collimated beams that may be brought to two focal spots in the plane of a diffraction limited lens. This lens performs a Fourier transform function, the splitting face of the prism being the same optical distance behind the back principal plane of the lens as is the focus of the two spots in front of the front principal plane. In fact, this distance should be slightly
modified to produce parallel focal zones in the test space. Converging beams at the first focus permit the relay lenses, whose longitudinal magnification would otherwise create diverging focal regions in the test space, to render these parallel.

The imaged spots are relayed into the test space by back-to-back diffraction limited collimator lenses. The spot size is determined by the diffraction limited image of the Gaussian input laser beam after magnification by the three collimator lenses. The separation is controlled by the divergence of the splitter prism, the focal length of the focusing lens and the magnification of the two output lenses. The spot diameter to separation ratio is, thus, easily designed to be an optimum value.

If the spot separation is \( s \), the beam splitter divergence is \( d \), the input laser beam diameter is \( D \), the half width at the Gaussian waist is \( r \), the focusing lens and internal and external lenses of focal lengths \( f_1 \), \( f_2 \), and \( f_3 \) respectively, we may write

\[
\frac{s}{2r} = \frac{d f_1 f_2 f_3}{2 f_1 f_2 f_3}
\]

and

\[
r = \frac{2}{\pi} (xf_1/D) (f_3/f_2) / r
\]

giving a spot separation to diameter ratio of

\[
\frac{s}{2r} = \frac{d D}{4}\frac{f_1 f_2}{f_3}
\]

4.2.2.2 Stops - The accurate design of stops is extremely critical for LTA systems. Their function is to reject light from all sources except that scattered from moving particles in the sampling volume. Although a simple pinhole at the conjugate image of each sampling region would seem adequate, several potential sources of flare must be considered. Multiple reflection from within the optical system, even with well-coated optics, are a source of scattered light. These may be reduced by double stopping where the image stops are themselves reimaged on another stop. This is effective in reducing second and higher order scatter sources, and has advantages not only in improving the rejection of flare generated from within the optical system, but also from multiple scatter in any test situation and beam walkoff sources in windows. Typical rejection factors of twelve orders of magnitude have been achieved.

In addition to the field stops discussed so far, two other types of stops are necessary to the system. An aperture stop is automatically provided by the outer edge of the final lens and by the edge of the turning mirror that separates outgoing from incoming light. It is necessary to extend the stopped area of the turning mirror to prevent diffraction flare from its edges. This is conveniently done by two more stops, one of which restricts the input beam to an area slightly smaller than the turning mirror, and one of which occludes an area slightly larger than the turning mirror from behind.

Flare may be seen by one photomultiplier that has arisen from light in the other channel. This may be prevented by stopping part of the received area. This results in a reduction in optical efficiency of the machine, and this stop is only necessary when wall proximity is important. This is executed as an extension to the stop immediately adjacent to the turning mirror.

4.2.2.3 Spot Intensity and Aberrations - To achieve the maximum benefit of the LTA principle requires that the system be nearly diffraction limited. Because of the small
field, the primary aberration is spherical. Projected spot quality and receiver efficiency are both impaired by normal and more by oblique incidence through a thick window. Typically, up to 5 cm thick windows with less than 10 degree incidence do not cause serious deterioration with F/4 systems. Curved windows, like large angles of incidence, introduce astigmatism.

The photon flux $F_i$ through the sampling volume is

$$F_i = \frac{P}{hv r^2} \text{photons/sec/area}$$

where $P$ is the laser power, $r$ is the radius, and $hv$ is the energy per optical photon.

The photon flux, scattered by a particle of radius $a$, during its transit of the peak, is $F_i q(a) = F q(a)$, where $q$ is the scattering efficiency. Hence, the total number of photons scattered during the transit, $G_s$, will be the flux multiplied by the transit time, $2r/V$, hence

$$G_s = 2a^2Pq/hv rV$$

This is necessarily approximate, because no allowance is made for which part of the illuminated region is intercepted by the particle, nor the profile function, nor the collection and detection efficiency for scattered photons.

4.2.2.4 Spot Rotation - Rotating the orientation of the two spots is required to determine flow direction. This can be done by 1) rotating the beam splitter and the stops, 2) using an image rotator prior to the output lens, and 3) rotating the whole machine.

The last method is simple but mechanically difficult. The first two lead to a more compact design. The first has more moving parts with opportunity for less than perfect alignment. The second is efficient and rather expensive. It is usual to arrange for computer control of rotation with corrective feedback of the angular position attained. Rotational precisions of 0.1 degree are easily attainable.

4.2.2.5 Detection - The detection of a transit event is straightforward, but, as the system becomes photon limited, other factors come into play. In an experiment without additional background light, where only light scattered by isolated particles is detected, two to four photons per spot are required for an LTA. This could actually be reduced to one per spot if the spots are far enough apart and there are no other photon sources.

In most experiments, the peak photon rate is almost always too high for quantum resolution, and analog filtering represents an excellent method of discriminating events from nonevents, where an event is associated with a group of photons from a single particle. A separate filter, that makes an estimate of time of occurrence of the highest instantaneous photon flux, may generate a digital signal to be correlated with a similar signal from the second spot.

4.2.3 Signal Processing and Interpretation

4.2.3.1 Broadening - Broadening of the data from an LTA is a manifestation of the inability to know precisely when the SC crossed the centers of the two spots. Broadening decreases as the spot diameter to separation decreases. Signal-to-noise ratio is critical here. When a photon limited signal occurs, a positional error for any one event can be almost as high as the spot diameter.

4.2.3.2 Data Rate - Data rate is proportional to the swept volume. This is the cross sectional area normal to the flow multiplied by the velocity. Natural aerosol distributions exhibit a number density approximately proportional to the inverse fourth power of radius, so, in an unseeded flow, the data rate can be vastly increased by extending the detection range to smaller SCs.

4.2.3.3 Time Measurement - Direct timing and correlation are used to measure the time intervals. In sparse particle conditions, where the transit time from one spot to the other is short compared with the mean interval between particles, there appears at first to be no significant advantage of one method over the other. Where particles occur more frequently, the efficiency of timing falls rapidly compared with correlation. In all cases, it is much more satisfactory to use two photomultipliers. With one PMT, timing can only stop on the next consecutive event, and this yields a background that must be compensated and is difficult to model. Autocorrelation yields a flat and subtractable background that is four times greater than that derived by cross-correlation of the signals from two photomultipliers.

4.2.3.4 Data Analysis - The signal from the two photodetectors are cross-correlated to produce a transit time probability curve. Because the background level is constant, it can be subtracted from the curve. The resulting transit time probability histogram must be converted to a velocity probability distribution. A velocity bias must be removed in the conversion. As with all LV systems, the effects of SC bias must be considered. The two-spot system measures magnitude and angle over an angular range governed by the spacing and diameters of the spots, typically the order of one degree. A curve of the type illustrated in Figure 4.30 is generated for each angular setting of the two spots. For a full angular sweep, a curve such as that shown in Figure 4.39 results.
Figure 4.38 (a) Transit time correlograms, (b) Velocity probability curve from (a) and (c) Bias introduced by multichannel analysis.

Figure 4.39 Velocity/probability plot from laser anemometer.

4.2.4 Comparison of LTA to LDV

Applying the equations presented above and those developed earlier for LDV, it is straightforward to compare the properties of the two system types. For particles of radius, \( a \), and sample volume radii of \( r_D \) and \( r_T \) for an LDV and an LTA respectively, assume that the same laser power is used for each system. Some comparative parameters for the systems are shown in Table 4.1.86

Table 4.1
RATIOS OF LTA VERSUS LDV FOR VARIOUS PROPERTIES OF EQUIVALENT SYSTEMS

<table>
<thead>
<tr>
<th>Ratio LTA/LDV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of photons scattered in transit</td>
<td>( \frac{2r_D}{r_T} )</td>
</tr>
<tr>
<td>Peak photon rate detected</td>
<td>( 2 \left( \frac{r_D}{r_T} \right)^2 )</td>
</tr>
<tr>
<td>Sample volume length</td>
<td>( \frac{r_T}{r_D} )</td>
</tr>
<tr>
<td>Data rate for monodispersion</td>
<td>( \left( \frac{r_T}{r_D} \right)^2 )</td>
</tr>
<tr>
<td>Data rate for ( n(a) a^{-4} ) (( a_D, a_T ) = minimum detectable radius)</td>
<td>( \left( \frac{r_T}{r_D} \right)^2 \left( \frac{a_T}{a_D} \right)^3 )</td>
</tr>
<tr>
<td>Data rate for ( N(a) ) (with allowance ( q(a) ) and retrieval efficiency)</td>
<td>( \left( \frac{3}{4} \right)^{3/4} \left( \frac{r_T}{r_D} \right)^{5/4} )</td>
</tr>
</tbody>
</table>
A typical ratio of probe volume radii for high-speed flow \( r_p/r_T \) is ten to twenty. Using this with the quantities in the above table, a number of generalisations can be made.

(1) The number of photons scattered for a given SC is larger by over an order of magnitude for the LTA. Therefore, the LTA can detect smaller particles. This tends to increase the data rate, although not much since the scattering cross section increases as the fourth power of the SC radius.

(2) The sample volume for an LDV is much larger than for an LTA. This can make the data rate for the LDV higher than that for the LTA unless the LTA is able to detect many more small particles than the LDV. When the flow is seeded with particles that can be detected by both systems, then the data rate for the LDV will be higher. The smaller sample volume length of the LTA allows measurements to be made closer to a surface than with an LDV. Typical ratio of probe volume radii for high-speed flow \( r_p/r_T \) is 10 to 20.

(3) The LTA is not easily used in highly turbulent or vortical flow.

(4) The LTA can be used in higher speed flows, since the spot separation can be adjusted to improve the accuracy of the transit time measurement. For the same reason, the speed measurement can be made more accurately.

(5) In nearly laminar flow, the flow direction can be measured more accurately with the LTA.

4.2.5 Applications

LTAs have been applied successfully in a variety of flow fields. In the following sections, some examples are given to illustrate the features of the technique. LTAs are commercially available from at least two companies.

4.2.5.1 Turbomachinery - The mapping of flow velocity between the compressor blades of turbomachines \( \text{Rolls-Royce}^{77} \) has been an important application of LTA. This is an extremely demanding application, because the measurement must be made in a backscatter mode, through small windows that quickly become dirty, close to both the windows and the blades, and with little or no seeding of the flow. All of these are conditions in which the LTA excels.

Scientists at Rolls-Royce \( ^{77} \) pioneered this application. Their system, illustrated in Figure 4.40, employs a two-watt argon laser focusing its energy into two 16-micrometer diameter spots separated by 160 micrometers. The optical assembly is contained in a box that is traversed by a three-dimensional traversing system. The measurements on rotating blade rows are made by gating one of the two photomultiplier tubes, so that it accepts data only when the sampling volume is in one selected part of the blade row. The gate is typically 5 or 10 microseconds long and is at a known time from a triggering signal produced by a blade. Alternately, it can be set to occur on every passage to obtain the average velocity at the same position in all passages. Typical velocities measured in this application range from a few m/sec to hundreds of m/sec.

Detailed velocity maps have been produced between the blades, resulting in the type of data illustrated in Figure 4.41.

4.2.5.2 Wind Tunnels - LTA is useful in wind tunnels, especially where only small SC are present and where the scattered light is too small for LDV. Also, in applications where measurement must be made close to models or windows, LTA excels. LTA is an excellent choice for measuring flow angularity. Even so, LDV has largely dominated wind tunnel applications because of versatility and ability to characterize turbulence.

4.2.6 Two-Line Anemometers

To overcome one of the principal limitations of LTA, researchers have developed systems that employ two lines instead of two spots. \( ^{43} \) The optics for such a system are illustrated in Figure 4.42. The unique properties of several optical components provide an efficient system. The two major lines of an argon laser are used to produce the two lines. The two lines are produced by separating the two colors angularly with a prism. Converting the two spots into two lines can be done by introducing astigmatism into the optical system. In the illustration, this is accomplished by passing the diverging light through tilted plates. The lines in this system are 20 by 100 micrometers in diameter. The collected, scattered light is separated into its two-color components by a dichroic reflector.

The two lines can also be generated by using cylindrical optics in the projecting system. Although somewhat more complicated, this produces a better sample volume. The astigmatic focus is complicated by the existence of a circle of least confusion, and a second focal region that contains a pair of lines oriented in a direction normal to the first.

The two-line LTA offers some of the advantages of both LDV and LTA. The system can handle flows of greater turbulence, but the benefits of rotating the two spots are lost. These kinds of systems are relatively new and their full benefits have not been widely demonstrated.
Figure 4.40 LTA System of Rolls Royce.

(a) Measurement Points Within a Blade Passage
(b) Derived Blade Relative Mach Contours

Figure 4.41 Laser anemometer results at 70% blade height in a research fan.
Figure 4.42 Scheme of the laboratory "two dashes" velocimeter.
CHAPTER 5
LASER SPECTROSCOPY AND SCATTERING METHODS

5.0 INTRODUCTION AND BACKGROUND

Light (photon) interaction with molecular matter is classified as either "elastic" (wherein the photons do not lose or gain energy) or "inelastic" (wherein the photons do exchange energy with the internal energy states of the matter). The classifications are illustrated in Figure 5.1.

INTERACTION BETWEEN LIGHT AND MATTER

<table>
<thead>
<tr>
<th>ELASTIC</th>
<th>INELASTIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIGHT IS SCATTERED INSTANTANEOUSLY</td>
<td>MODULATION OF THE INCIDENT LASER BEAM</td>
</tr>
<tr>
<td>WITHOUT EXCHANGING ENERGY WITH THE INTERNAL</td>
<td>BY THE INTERNAL ENERGY STRUCTURE</td>
</tr>
<tr>
<td>STATES OF THE GAS MOLECULES</td>
<td>OF SPECIFIC MOLECULES</td>
</tr>
</tbody>
</table>

RAYLEIGH SCATTERING | MIE SCATTERING | ABSORPTION | FLUORESCENCE | RAMAN EFFECT

a. Overall Classification

RAMAN EFFECT

<table>
<thead>
<tr>
<th>LINEAR</th>
<th>NONLINEAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPONTANEOUS RAMAN SCATTERING (CORS)</td>
<td>COHERENT</td>
</tr>
<tr>
<td>(CORS)</td>
<td>INCOHERENT</td>
</tr>
<tr>
<td>VIBRATIONAL</td>
<td>ROTATIONAL</td>
</tr>
<tr>
<td></td>
<td>ROTO-VIBRATIONAL</td>
</tr>
</tbody>
</table>
| Stokes $\omega_s < \omega_L$ | Anti-Stokes $\omega_s > \omega_L$

b. Raman Interaction

RAMAN ACRONYMS

CORS Conventional Old-Fashioned Ordinary Raman Scattering
CARS Coherent Anti-Stokes Raman Scattering
IRS Inverse Raman Scattering
CSRS Coherent Stokes Raman Spectroscopy (pronounced scissors)
RIKES Raman Induced Kerr Effect Spectroscopy
HORES Higher Order Stokes Effect Scattering
PARS Photo-Acoustic Raman Scattering
CARE Coherent Anti-Stokes Raman Ellipsometry
The elastic scattering process is essentially instantaneous. One can think of the photon striking the molecule and rebounding with no energy loss like a hard sphere collision. Electromagnetic theory models the interaction as the acceleration of the bound electrons in the molecules by the changing electric field of the impinging radiation and the electrons radiate at the same frequency as the field which causes their acceleration. Even though no energy exchange occurs, the interaction can be used for diagnostics because changes are effected in the distribution of the radiation in space as well as the polarization state of the radiation. Also, if the molecule is in motion, the frequency of the radiation is Doppler shifted by the scattering center by an amount that is directly proportional to its velocity.

Each molecule has a scattering cross section that determines the amount of light scattered from impinging radiation as well as how the scattered radiation is distributed in space. The cross section depends on the electrical properties and the size of the scatter centers. The total amount of radiation scattered from the impinging wave is directly proportional to the number of molecules and thus, the number density.
The diagnostics process can be further enhanced by illuminating the volume of interest with well defined distributions of light, such as one or more spots or plane sheets. Both transmitting the light to the region of interest as well as collecting the scattered light in a carefully restricted way with specially designed apertures and lenses can improve the signal to noise ratio and help to spatially resolve the place of measurement.

As scatter centers pass from one point to another in a well defined light field, the time and space dependence of the scattered light can be used to determine velocity and turbulence, as described in Chapter 4.

A photon can interact with matter inelastically in the following five different ways (Figure 5.1c).

1. Absorption - The photon is absorbed entirely through an energy exchange process in which its energy is absorbed, forcing the molecule into a higher energy state. This can be an electronic energy state or a molecular rotational or vibrational state or a combination of all three. The change of energy of the molecule equals the energy of the photon.

   Subsequently, the molecule can decay spontaneously to a lower energy state while emitting a photon whose energy equals the change in energy of the molecule. Spontaneous emission occurs after about $10^{-8}$ seconds for most molecules. Some states are much longer lived, lasting even seconds. These are called the fluorescent states.

2. Stimulated Emission - The striking photon stimulates (or drives) an excited molecule down to a lower energy state resulting in a decrease in energy equal to the photon energy. This transpires three orders of magnitude faster than spontaneous emission. Then the striking photon and the emitted photon will have the same frequency and leave the molecule in phase with each other.

3. Raman-Stokes Scattering - A photon strikes a molecule, adds part of its energy, and drives it to a higher level. To conserve energy, a photon must be emitted to carry away the remaining energy. The scattered photon has lower energy than the striking photon (and, therefore, lower frequency). The probability (or cross section) for this type of interaction is about three orders of magnitude lower than that for elastic scattering.

4. Raman-Anti-Stokes Scattering - An already excited molecule is driven to a lower level by a striking photon causing an energy decrease in the molecule that (unlike the stimulated emission described above) is not related to the energy of the striking photon (and, therefore, lower frequency). The probability (or cross section) for this interaction is even less than that for the Stokes interaction.

5. Stimulated Raman Scattering - The photon strikes a molecule that has been excited to a virtual state and stimulates (drives) it to a lower state with an energy decrease equal to the energy of the photon. This can take place only in conjunction with another photon that is undergoing Raman scattering. The photons that leave the interaction have a distinct phase relationship with respect to each other.

This chapter covers:

1. elastic scattering from molecules (Rayleigh Scattering) to measure density, temperature and velocity, and
2. inelastic scattering (absorption/emission, fluorescence, and Raman spectroscopy) to measure temperature, species, velocity, and density.

5.1 Rayleigh Scattering

The measurable parameters using Rayleigh scattering are temperature, density, and velocity.

5.1.1 Principle of Operation

Figure 5.2 illustrates a typical Rayleigh scattering instrument for the measurement of the above parameters. A pulsed laser is focused to produce a high intensity illumination of a small volume in the gas. A large aperture, high resolution imaging system further restricts the collection of scattered light to a small volume in space. The total amount of scattered light is proportional to the number of molecules illuminated and, therefore, is proportional to gas number density.

The linewidth of the scattered illumination is broadened by the Doppler shift impressed on the scattered light by the motion of the molecules and, therefore, is proportional to $(\text{temperature})^{1/2}$. The center frequency of the scattered light is Doppler shifted by the bulk motion of the gas by an amount that is proportional to the gas velocity.
The Rayleigh scattering cross section for ambient air is large compared to the cross section for inelastic scattering, but it is still so small that high power lasers must be used to produce a measurable amount of scattered light. Consequently, a pulsed laser must be used. The cross section varies directly with the minus fourth power of the wavelength, so much gain is to be had by using short wavelength lasers. This technique has flourished especially with the emergence of ultraviolet excimer lasers that can provide short wavelength light at intensities never before available.

Perhaps the greatest attraction of all for Rayleigh scattering is the prospect of measuring temperature, density, and molecular velocity all simultaneously (Refs. 1 and 2). The essential feature of Rayleigh scattering is that the scattered signal is not shifted in frequency with respect to the incident beam (except for a possible Doppler shift).

5.1.2 Analytical Description

The intensity of Rayleigh scattered light, \( I_r \), from an incident beam of intensity, \( I \), can be written as:

\[
I_r = I_L \sigma N a
\]

where \( N \) is the gas number density, \( a \), the solid angle of collection, \( \sigma \) the probe volume length in the laser beam direction, and \( \sigma \) the total scattering cross section. If the gas constituents are known, the total scattering cross section may be calculated from the scattering cross section of the individual constituents such that

\[
\sigma = \sum_{i=1}^{N} x_i \sigma_i
\]

where \( x_i \) is the mass fraction of constituent \( i \) and \( \sigma_i \) is its scattering cross section. Individual Rayleigh scattering cross sections can be accurately predicted using Mie theory, and are functions of the wavelength, complex index of refraction and the depolarization factor.
Equations 5.1 and 5.2 can be used to determine the number density from the measured $I_R$ and the temperature can be deduced from the spectrum of the scattered light. The spectral distribution is formed by the Doppler shifts of the light scattered from individual molecules in thermal motion. Namely,

$$I_\lambda = \frac{I_R L}{\lambda_0^2} \exp \left(-\frac{\Delta \lambda}{\lambda_0}\right) \quad (5.3)$$

where

$$L^2 = \frac{H c^2}{8kT} \lambda_0^2 \sin^2 \left(\theta/2\right) \quad (5.4)$$

and $c$ is the speed of light, $k$ is Boltzmann's constant, $H$ is the molecular weight, $T$ the translational temperature and $\Delta \lambda = \lambda - \lambda_0$ is the shift from the incident wavelength. Take, for example, nitrogen molecules, a laser wavelength of 488 nm, and backscattering ($\theta = 180$ degrees), then the full width half maximum will be related to the temperature through

$$\text{FWHM} = 0.0235 \frac{\Delta \lambda}{1000(4K)} \lambda \quad (5.5)$$

The ability of the technique to measure temperature depends on the laser line width. Typical laser linewidth at 300 nm is of the order of 0.001 $\lambda$, which is far below the FWHM calculated from Equation 5.5.

5.2 Absorption/Emission Spectroscopy

The measurable parameters using absorption/emission spectroscopy are temperature, pressure, species concentration and velocity.

5.2.1 Principle of Operation

Emission spectroscopy analyzes the spontaneous radiation spectrum from a sample volume as compared to absorption spectroscopy which measures the attenuation as a function of wavelength of radiation that is passed through the medium from a light probe (Figure 5.3). The temperature is inferred from the shape of the absorption or emission profile. The concentration or pressure is inferred from the amount of emission or absorption, and the species concentration is inferred from the overall absorption or emission spectrum.

![Figure 5.3. Species identification and temperature measurement by absorption spectroscopy](image)

In the case of absorption, the source may be either continuous, discrete, or ultra-narrow bandwidth in the form of a laser. This discussion will concentrate on absorption spectroscopy using laser light sources. With a laser source the spectral purity can be extremely high and the spectral linewidth can be less than the individual lines in the absorption spectrum. Therefore, one laser line can be tuned over an absorption line to measure the broadening of that line to infer temperature.
The measurement is an integrated effect and like interferometry, the point function must be determined by solving the integral. If the distribution of the measured parameter is constant, the solution is simple. If the distribution is axisymmetric the solution is the so-called Abel inversion. If the distribution is three-dimensional, then a tomographic analysis such as described in Chapter 2 is required to produce the point measurement.

5.2.2 Analytical Description

The intensity of a beam of light passing through a medium along s changes in accordance with the following equation:

\[ \frac{di}{ds} = -k_v (i_v - B_v) \]  \hspace{1cm} (5.6)

where \( i_v \) is the intensity in watts/steradian/cm², \( B_v \) is the blackbody function, and \( k_v \) is the absorption coefficient. The solution to this equation contains two terms, one describing the absorption of the incident radiation and the second describing the radiation added along the path from molecular emission.

\[ i_v = i_v^0 \exp \left( \int k_v \, ds \right) + \int k_v \, B_v \exp \left( \int -k_v \, ds \right) \, ds \]  \hspace{1cm} (5.7)

The second term depends on the blackbody function times the absorption coefficient integrated along the pathlength attenuated by the intervening gas between the observer and the emitting volume. The absorption coefficient for a single transition, \( k_v \), in units of inverse length is given by

\[ k_v = \frac{c^2 f A_v}{8 \pi^2 \hbar^2} \left[ \frac{N}{N_1 - N_2} \right] f(v) \]  \hspace{1cm} (5.8)

where \( A \) is the Einstein A coefficient, \( f \) the normalized lineshape function, \( N \) the state number density, and \( g \) the degeneracy of the energy level.

The Einstein coefficient indicates the absolute probability of a transition from one energy state to another and is determined by the quantum mechanics of the two states.

The lineshape function depends upon several mechanisms. For relatively low pressures, Doppler broadening dominates and the shape parameter is a Gaussian with temperature as the primary variable. If pressure is high enough, the broadening mechanism becomes Lorentzian in shape with the collision frequency the dominant factor. Collision frequency depends on both pressure and temperature.

The Doppler width is given by

\[ b_d = \sqrt{\frac{2 kT \pi \nu}{\hbar^2}} \]  \hspace{1cm} (5.9)

and the collisional width is given by

\[ b_c = b_{c0} \left( \frac{T}{T_0} \right)^{1/2} \]  \hspace{1cm} (5.10)

where \( b_{c0} \) is the measured width at a standard temperature and pressure.

If Doppler and collisional mechanisms are of the same order, then the Voight (mixed lineshape) function must be used (see, for example, Refs. 3 and 4).

Absorption spectroscopy has not yet been widely used in flow diagnostics; however, the availability of usable hardware and the need for diagnostics of unseeded flows is likely to change this in the near future. The strong oxygen absorption band in the 170-190 nm region can be used to measure temperature with an excimer laser as a light source. The technique, in principle, is rather simple, covers a wide dynamic range and requires only small windows.

5.3 Laser Induced Fluorescence

The measurable parameters using laser induced fluorescence are number density of the fluorescing species and temperature (from which pressure can be derived from the equation of state) and velocity.

5.3.1 Principle of Operation

A tunable laser is tuned to resonantly excite a particular transition of a species in the test volume. The species may be naturally occurring or perhaps a more convenient species introduced into the flow as a seeding material. Fluorescence is a spontaneous radiation that arises from the decay of the excited molecules. A typical experimental arrangement for laser induced fluorescence measurement is shown in Figure 5.4. The fluorescence signal can be used in the following ways:

1. The population of the excited state, which is proportional to the amount of radiation can be related to the total number density of the species.
2. By measuring the population of two states near the ground state, the population of the ground state and the temperature can be determined from the Boltzmann distribution.

3. The narrow linewidth laser can be tuned over the absorption line and the fluorescence spectrum compared with that of a reference stationary cell (see Reference 5). The frequency shift of the center frequency is the result of the Doppler shift caused by moving molecules. The broadening of the spectrum can provide temperature and pressure if the broadening parameters are known. The requirement of a reference cell can be obviated by using counter-propagating laser beams. This provides a Doppler upshift from one of the beams and a Doppler downshift from the other, and the average of the two is the equivalent of the stationary velocity.

![Figure 5.4 Planar laser induced fluorescence (PLIF) measurement of gas density.](image)

5.3.2 Analytical Description

An approach for analyzing the fluorescence signal is described by Hiller, Reference 6. The laser frequency is tuned to the linear region of the fluorescence intensity versus wavelength curve. The fluorescence from two opposing beams is Doppler shifted up and down by the velocity of the gas, and the amount of the intensity change corresponding to the Doppler shift is (see Figure 5.5)

\[
\Delta S = \frac{S - S_b}{2}
\]  

(5.11)

![Figure 5.5. Fluorescence intensity versus frequency for static absorption line (heavy line) and for blue-shifted (laser beam in same direction as velocity component) and red-shifted cases. The curves are drawn for one specific velocity \(u\). The laser frequency is \(\nu_L\). The approximations assume that \(S_L\) and \(S_B\) are located on the linear part of the line. (After Hanson, Ref. 7)](image)
The frequency shift is
\[ \Delta f = \Delta S \left( \frac{\Delta S}{\nu} \right)^{-1} \]  
(5.12)

The derivative must be determined from the calculated or measured specular distribution function.

In the work reported in Reference 6, three components of velocity were measured by probing the gas from three directions with four laser sheets. The measurement was done in 250 ms.

Recent implementations of this method employ laser sheets so that the velocity distribution over an entire plane can be measured. The sheet of light is imaged onto a photodiode array camera after having been intensified by an image intensifier. The laser beams from the different directions can be recorded successively with a full matrix of intensities being recorded for each. Then Equation 5.12 is solved for Doppler shift for each point in the sheet.

5.3.3 Special Considerations

The limiting factors in the accuracy of the velocity measurement are:

1. Lifetime Broadening
2. Collision Broadening
3. Thermal Broadening
4. Laser Broadening
5. Turbulence Broadening

The typical errors cited by researchers is about 5 m/s. This technique is more applicable to high speed flows, and has been demonstrated in subsonic jets at about 50 m/s (Ref. 7) and for hypersonic flows up to about 1700 m/s (Ref. 8).

To be used at low speeds, the gas pressure must also be small to produce narrow lines from which the Doppler shift can be extracted. The highest practical pressure is about one atmosphere (Ref. 9). Other limitations are caused by the imperfections or dirt on the test cell windows that will affect the fluorescence intensity. Spectral analysis of the fluorescence circumvents this problem at the expense of a longer processing time.

5.4 Raman Spectroscopy

The measurable parameters are species concentration, vibrational and/or rotational temperature of the particular species, and flow velocity. Translational temperature can be estimated with considerably less accuracy from the Doppler broadening of the scattered light. The technique is typically useful for concentrations of about 1 part per million.

The variations of Raman spectroscopy are many (see Figure 5.1) and cannot all be discussed in this work. However, the most important are Conventional Raman (CDRSR), Inverse Raman (IRS) and Coherent Raman Spectroscopy (CARS and CRARS) and a brief discussion is included.

5.4.1 Principle of Operation

About one out of every thousand times a photon strikes a molecule it loses a fraction of its energy to the molecule and alters the internal energy states of the molecule, and the emerging photon has a longer wavelength (typically tens of angstroms) from the original photon. This is called Raman-Stokes scattering. The observed frequency shift depends upon which energy states were involved in the interaction, so what is observed is actually a spectrum of frequencies that are characteristic of the molecule. Every molecule has a unique Raman spectrum. This provides a fingerprint of the molecule and is, therefore, a diagnostics for the type of species present in the sample.

How the intensities are distributed between the frequencies in the Raman spectrum depends on the original distribution of energy states in the sample, therefore, providing a measure of the temperature associated with the particular state, such as vibrational or rotational temperature.

In a related but less probable interaction a photon colliding with a molecule can emerge from the interaction with greater energy (and higher frequency) by leaving the molecule in a lower energy state. The scattering process is called Raman anti-Stokes. The same fingerprinting characteristic applies, and the ratio of the Stokes to anti-Stokes scattering cross section provides a measure of the initial distribution between energy states and, therefore, can be used to measure temperature.

In any of the cases, the flow velocity of the molecules can be determined through the Doppler shift of the scattered radiation. A calibration cell is used to provide the spectrum with no Doppler shift and that is compared with the spectrum of the light scattered from the volume under measurement.
5.4.2 Technique Description

Raman spectroscopy is relatively simple experimentally. A laser beam is concentrated in a region of interest and the scattered light is collected and analyzed (see Figure 5.6). The largest amount of light emerging from the sample volume arises from Rayleigh or Mie scattering, or fluorescence. However, the Raman scattering occurs at a different wavelength and can be separated from the composite by using narrow-band filters, or a monochromator. Then the intensity of the light is measured with a photodetector or recorded on film.

![Figure 5.6. Species identification by Raman Scattering.](image)

The complete spectral analysis of the signal requires a monochromator or spectrograph, but in special cases the measurement is much simpler. For example, if the constituents are known, then their concentration can be measured simply by measuring the amount of light existing in certain narrow spectral bands. Alternately, the distribution of a species in space can be observed by illuminating the volume with a laser sheet, then recording through a narrow line filter centered about one of the Raman spectral components of the species. This is called a Raman graph. The signal is extremely weak, so a combination of image intensifiers, multiple exposures, and high power lasers is required.

When the laser line approaches resonance with a state in one of the molecular species, Raman scattering is enhanced. The transition and, indeed, the distinction between resonance Raman and resonance fluorescence is not completely understood, but Raman scattering is defined as that process that is almost instantaneous, involving virtual states while fluorescence is a relatively time-consuming process (typically microseconds) involving real states.

The structure of the Raman spectrum can be extremely complex, even for simple molecules, because all of the allowable transitions in the molecule can become involved. The impinging photon can drive the molecule into any of the allowed final states as long as energy is conserved. In quantum mechanics, the allowable changes in quantum number in an energy transition are zero and plus or minus one for vibration and zero or plus or minus two for rotation. So for a given vibrational state transition, for example, a change of vibrational quantum number zero to one, the spectrum contains three branches. The excitation to higher rotational quantum numbers produces what is called the S branch, while excitation to the lower quantum numbers produces the O branch, and no change in rotational number produces the Q branch.
5.5 Coherent Raman Spectroscopy

The measurable parameters using CARS, CSRS, and IRS are the same as with Raman spectroscopy, that is, species, and rotational/vibrational temperature and flow velocity.

5.5.1 Principle of Operation

The difficulty in using conventional Raman spectroscopy arises from the small Raman cross section and also from the homogeneous scattering in all directions, making the collection of scattered light inefficient. Coherent Raman solves this problem by stimulating the process with a tunable probe laser which forces the molecular transition to occur with a fixed phase relationship to the pump and stimulus beams. Figure 5.7 illustrates a CARS experiment.

![Energy Level Diagram](image)

(b) Physical Process

![Experimental Set Up](image)

Figure 5.7. Description of a CARS experiment

1. The impinging photon from a pump laser raises the energy state of the molecule to a virtual level, which is $h_{vp}$ above the initial state.
2. The stimulus or probe laser is tuned to stimulate a downward transition to a state below the virtual state by an amount $h_{vs}$.
3. A second photon strikes the molecule adding energy $h_{vp}$ to the molecule raising it to a still higher virtual state.
4. The molecule decays to the initial state emitting a photon of energy $2h_{vp} - h_{vs}$.
5. The emitted photon leaves in a preferred direction, defined by the directions of the input beams. This is a result of a constant phase relationship between the photons.
6. The probe laser can be tuned over the entire Raman spectrum allowing a plot of the spectrum with a resolution that depends only on the linewidth of the probe laser.
7. In CSRS the process involves the stimulated Stokes process.
8. In IRS the energy removed from the probe beam by the stimulation process is measured. The process is called "inverse" since a decrease in intensity is observed.
5.6 Application Examples

5.6.1 Rayleigh Scattering for Measurement of Density in the Shuttle Reentry Path

Rayleigh scattering has not yet been applied routinely in aerodynamics. An application planned by McKinsey (Ref. 10) for measurement of the density along the flight path and the flow field near the space shuttle during its reentry provides a good example of the potential of the technique. The planned system is illustrated in Figure 5.8. The recommended system will employ an ArF laser operating at the 193 nanometer wavelength. A pulsed output energy of less than 1 millijoule was shown to be sufficient to provide an adequate return signal. (Commercially available lasers now exceed 200 millijoules at pulse repetition rates in excess of 200 Hz.) The Cassegrain telescope has a collecting aperture of 23 cm. The photomultiplier is CsTe on fused silica having a quantum efficiency of 15% and a maximum gain of $2 \times 10^4$. A 22 nm narrow-band filter removes ambient light.

The system will provide a Rayleigh return from oxygen and nitrogen molecules in a column that extends out to about one kilometer from the shuttle. The interaction effect allowing contribution from all of the molecules in the long column is one of the keys to the success of the application. The cross section of air used for the computations was about $3.5 \times 10^{-26}$ cm$^2$. (After McKenzie, Ref. 10)

Computations of the signal-to-noise ratio show that such a system is capable of providing direct measurements of ambient density with an uncertainty of less than 1% over an altitude range from 50 to 90 km with a spatial resolution of 1 km.

McKinsey has also produced a conceptual design of an instrument that could spatially resolve the density profile inside the shock wave of the shuttle and could actually locate the shock. Such a system must time resolve the return signal and associate the intensity of the return signal at a known time with a distance from the shuttle. This is a much more demanding task both on the laser and on the detection system, since one must detect scattered power and not total integrated energy as in the above application. Therefore, a much more powerful laser is required and a detection system that has a frequency response in excess of 1 GHz for a resolution of about .3 meters.

5.6.2 Laser Induced Fluorescence Measurements of Oxygen in Supersonic Flows

Temperature measurements have been made in supersonic flows below 300 K using laser induced fluorescence on NO seeded into a nitrogen flow (Ref. 11) and iodine seeded into an air flow (Ref. 12). Velocity and pressure fields in supersonic nozzle flows have been determined using the Doppler shift of molecular absorption lines for iodine seeded into a nitrogen flow (Ref. 13). The use of seeding materials in flows is rarely desirable or easy to maintain. Unfortunately, some of the best seeding materials for fluorescence are highly toxic and unsafe to use.

An extremely attractive candidate for LIF measurements would be molecular oxygen because of its natural presence in most flows of interest. Reference 14 describes a study in which oxygen was used as the fluorescing material. Available excimer lasers operating far in the ultraviolet make such measurements possible. The availability of the narrow-band laser source also makes LIF Doppler-shift velocity measurements possible. If a component of the velocity lies in the direction of the exciting radiation, the molecules absorb at a frequency that is Doppler-shifted from the laser frequency. Since the place at which the laser falls on the absorption lineshape function
4. The detector senses light changes in the probe beam caused velocimetry although the technique with one instrument. RDV possesses significant advantages over Rayleigh Doppler solved. Also attractive is the prospect of measuring temperature, density, and velocity demonstrated to date but some possibilities of using two exciting wavelengths with broadband collection or a single wavelength was suggested. Where multiple parameters are varying, the possibility of using two exciting wavelengths with UV-grade fused silica.

The sensor was placed at 90 degrees to the sheet and viewed through a KBr high-pass filter. The sensor comprised an ITT Model 4144 dual-plate image intensifier with an 8-20 cathode. Mounted to the image intensifier was a Reticon Model MC520 100 x 100 element array camera using fiber-optic bundle coupling. The data from the image array is then fed to a computer which processes and displays the data in two-dimensional formats.

The resulting oxygen fluorescence images demonstrated velocity, temperature, and pressure sensitivity. Separating the parameters velocity, temperature, and pressure is not a trivial process. The recommended procedure emerging from this study is to use an absorption line that is insensitive to temperature to measure pressure. The collection is done at 90 degrees where there is no Doppler shift to isolate the measurement from velocity dependence. To measure the temperature, a line that is highly dependent on temperature can be selected. Finally, velocity measurement must be done at an angle different from 90 degrees with respect to the illumination.

In reacting flows where the oxygen concentration changes, the measurement can be used to determine oxygen number density. Where multiple parameters are varying, the possibility of using two exciting wavelengths with broadband collection or a single wavelength with two detection channels was suggested.

5.6.3 Wind Tunnel Applications of Coherent Raman Scattering

A limited number of applications of Raman scattering in wind tunnels have been demonstrated to date but some of these are extremely promising. Raman Doppler velocimetry (RDV) is important since the particle lag problem in standard LDV is solved. Also attractive is the prospect of measuring temperature, density, and velocities with one instrument. RDV possesses significant advantages over Rayleigh Doppler velocimetry although the technique is much more complicated.

RDV has been demonstrated in a variety of forms. For example, CSRS (Ref. 15), IRS (Ref. 16), IRS (Ref. 17), stimulated Raman gain spectroscopy (SRGS) (Ref. 18) have all been demonstrated in wind tunnel environments. High resolution CARS and CSRS have been proposed as a method for measuring turbulence (Ref. 19).

Exemplar applications of Raman techniques have been demonstrated in the Unitary wind tunnel at NASA Langley Research Center, illustrating an important utility in high-speed flows (Ref. 20, 21). Specifically, IRS was used to measure temperature, density, and velocity. Figure 5.9 shows the somewhat unique experimental setup, which incorporates some finesse necessary for successful operation in a wind tunnel environment. The following are key features of the setup:

1. Molecules in the sample volume are illuminated by focused beams at two different wavelengths. One, the probe beam, is produced by a narrow-band (< 10 MHz linewidth) CW argon laser operating at a power of 0.5 watts at 514.5 nm wavelength. The other (the pump beam) is produced by an argon-laser-pumped dye laser whose output is amplified by a pulsed dye amplifier producing a 5 nsec, 7 mJ pulse at 584.6 nm wavelength. The pulsed dye amplifier is pumped by a Q-switched, frequency-doubled YAG laser.

2. After the beams have passed through the sample volume and to the other side of the tunnel, the pump beam is removed by a dichroic filter and dumped.

3. The probe beam is returned by a retroreflector to focus again (from the opposite side) on the molecules in the sample volume.

4. The detector senses light changes in the probe beam caused by two different types of interaction. One of the interactions is with the two beams propagating in the same direction (when the photons enter the sample volume for the first pass). The second interaction is with the two beams counterpropagating (when the probe beam returns to the sample volume from the retroreflector).

5. The intensity change in the probe beam is measured while tuning it over the Raman spectrum. This intensity decreases most when the energy of the probe beam photons are properly matched with the energies of the pump beam photons to allow a resonance with the final energy state of the molecule. The probe beam drives the molecular energy up to a virtual energy state and the pump beam stimulates it down to an allowed state. The emitted photon has energy which is the difference in the probe and pump photon energies. This is called the IRS resonance.
6. The Doppler shift effects caused by the motions of the molecules is significantly different for the two types of interactions (copropagation and counterpropagation), and these produce two distinct peaks in the measured spectrum. The measurement of the separation of the two peaks provides an accurate measure of velocity. The measurement of the difference in broadening of the two peaks provides a measure of temperature and pressure.

\[
\Delta \nu = \frac{2 \nu_s \cos \phi}{\lambda_p (1 + \alpha)}
\]

For molecules moving with a velocity \( V \), an illuminating wavefront is Doppler shifted by \( (k_s - k_0) \cdot V \) (see Chapter 6), where \( k_s \) is the illuminating wavefront and \( k_0 \) is the observed, scattered wavefront and \( k = \omega/2\pi \). Both the pump beam, \( k_{pump} \), and the probe beam, \( k_{probe} \), are "seen" by the molecule as Doppler shifted wavefronts. Therefore, the wavelength at which a resonance is observed when tuning the wavelength of the probe beam depends on the velocity of the molecule. Without Doppler shift the resonance occurs when \( \nu_{probe} = \nu_{pump} + \Delta \) where \( \Delta \) is the associated Raman frequency. The Doppler shifted resonance occurs when

\[
\nu_{pump} = (\nu_{probe} - \Delta) + (\nu_{pump} - \nu_{probe}) \cdot V
\]

The difference between the resonant wavelengths for the copropagating beams interaction and counterpropagating beams interaction can be determined by applying Equation 5.13 to be

\[
W = 2 \nu_{probe} \cos \phi
\]

where \( \phi \) is the angle between the probe beam direction and the velocity vector. In this manner, velocity can be measured without resorting to an absolute frequency measurement. Likewise, pressure and temperature can be determined from frequency difference measurement.

Figure 5.10 is an actual plot, produced by Exton, et al for a Mach 2.5 free-stream flow, of the power loss from the probe laser during the 5 nsec pump beam periods as the wavelength is tuned for the copropagating and the counterpropagating interactions. The scan time required was about two minutes with a time constant of about 0.3 seconds. The velocity of the flow is computed from the peak separation. For this case, the frequency difference between the peaks was 1.00 GHz corresponding to a velocity of 602 m/sec (Mach 2.31).
The pressure and temperature are computed by measuring the two line breadths. The two line breadths were 0.33 GHz for the copropagating interaction and 1.98 GHz for the counterpropagating interaction. A family of theoretical curves of the Voigt line breadths (temperature and pressure broadening) can then be used to derive temperature and pressure from the two measured line breadths. The corresponding pressure and temperature for this case were 35 torr and 161 K respectively.

The authors show that the precision in this measurement was about 17% for temperature and pressure and 5% for velocity. With an ideal signal the precision of this method would be about 2% for temperature and pressure and 1/2% for velocity. The technique is shown analytically to be applicable in the hypersonic regime at pressure down to 10 torr. Below that pressure, the signal begins to deteriorate below detectivity without other processing. With other signal enhancement procedures, this technique can be expected to cover virtually the entire current interest range in high speed aerodynamics and will doubtless prove to be an extremely important diagnostic tool for modern aerodynamics.

Coherent Raman techniques have seen extensive application in combustion diagnostics. Eckbreth and associates (see additional reading) have refined CARS techniques to routine use in combustors and jet engine development, and other combustion laboratories have used CARS extensively. The complexity of the hardware as illustrated in Figure 9 discourages its use in large-scale wind tunnel tests although a wider applicability can be anticipated.

Raman methods provide the unique capability to monitor several species simultaneously at a point. This can be contrasted to absorption and fluorescence which monitor a single species for each wavelength, therefore, requiring wavelength tuning to cover several species.
APPENDIX A

FIGURE A1. REPRESENTATION OF PLANE WAVEFRONTS

FIGURE A2. REPRESENTATION OF PLANE WAVEFRONTS OF WAVELENGTH SLIGHTLY LARGER THAN THAT OF FIGURE A1.
FIGURE A3. REPRESENTATION OF SPHERICAL WAVES.
FIGURE A4 SPHERICAL WAVES OF SLIGHTLY LARGER WAVELENGTH THAN IN FIGURE A3.
APPENDIX B

REFERENCES


3.26 Kasahagi, T., Jones, W., Kasahagi, T., and Baum, H., "Application of High-Speed Two-Wavelength Holographic Interferometry to the Analysis of Radiative Ignition," The American Society of Mechanical Engineers.


4.30a Buchhave, P., 1973, "Biasing Errors in Individual Particle Measurements with the LDA-Counter Signal Processor," Proc. LDA-Symposium, Copenhagen, P.O. Box 70, DK-2740, Skovlund, Denmark.


APPENDIX C

ADDITIONAL REFERENCES

CHAPTER 1

CHAPTER 2

CHAPTER 3


Surget, J., "Holographic device for rectilinear surface discharge visualisation." Third International Symposium on Flow Visualization, Ann Arbor (USA), TP. ONERA n° 1983-104.


CHAPTER 4


Cherdron, W., Durst, F., and Whitehead, J. H., "Asymmetric flows and instabilities in symmetric ducts with sudden expansions".

Durao, D. F. G., Durst, F., and Firmino, F. C., "Velocity Characteristics of the Flow Around Cones".


Durst, F., and Krebs, H., "LDA-Optics Developments for Measurements in Internal Combustion Engines".

Durst, F., and Schmitt, F., "Joint Laser-Doppler/Laser-Induced Fluorescence Measurements in a Turbulent Jet".


McKillop, A. A., and Durst, F., "LDA Experiments of Separated Flow Behind a Circular Cylinder".


CHAPTER 5


McCullough, R. W., "Resonant Doppler Velocimeter," In AGARD Applications of Nonintrusive Instruments in Fluid Flow Research, N77-11221, Publication Date: May 1976.


Miles, R. B., "Resonant Doppler Velocimeter," In AGARD Applications of Nonintrusive Instruments in Fluid Flow Research, N77-11221, Publication Date: May 1976.


ACKNOWLEDGEMENTS

One of the extraordinary pleasures in the preparation of an AGARDograph is provided in the visits to the laboratories in NATO countries to work with the world's greatest scientists. I was humbled with the kindness, openness and earnest desire to contribute to the project experienced during these visits and I cherish the resulting friendships that are the real payoff for being an AGARDograph author. I cannot express enough appreciation to those of you who made my visits such a valuable experience. I can only hope that I have done you justice in this summary of your work.

A second extraordinary pleasure in the preparation of an AGARDograph is in the acknowledgement of those who contributed. For this task tells me that the sorting out and reporting is almost over and, I can now reflect on the experience of sharing in the work of those who made this possible. So many people contributed that I will quite possibly forget one or misspell a name. Please forgive me for that.

I owe much to several friends who gave up a considerable amount of their time to serve as focal points in their institutes, setting up relevant interviews, collecting and sending papers and figures, and reviewing parts of the manuscript. These included Dr. Franz Durst, University of Nurnberg at Erlangen, Germany, Andre Girard of the National Office of Aerospace Research (ONERA), at Chatillon, France, Dr. Hans Pfeifer, French-German Institute at St. Louis, France, Dr. David Jones and Nick Parker, at Rolls Royce, in Derby, England, Drs. Nick Phillips and John Tyer, at the University of Loughboro, England, Dr. Peter Bryanston-Cross, at Cambridge University, England, Dr. Michel Reithmueller, at the von Karman Institute, at Rhode-St.-Genese, Belgium, Mr. George Lee at NASA Ames Research Center, California, USA.

I received considerable advice, papers and useful conversation from Dr. George Havener, State University of New York, Binghamton, USA, Dr. George Selbert, Wright Patterson AF Base, Ohio, USA, Dr. Ron Hanson, Stanford University, California, USA.

I owe thanks to Spectron, A Titan Company, and to Dr. Chris Busch, President of Spectron, for the support throughout the project. Drs. John Abbiss, Arun Arunkumar, Cecil Hess, Anthony Smart of Spectron reviewed parts of the manuscript. Mrs. Jacqueline Johnson, Mrs. Mary Kittle, and Mrs. Marnice Ziegler typed the manuscript and many of the figures were drawn by Mr. Art Kittle.

Finally, I would like to thank Dr. Jean Ginoux, Director of the von Karman Institute, for helping provide the opportunity, and for acting as technical editor and Dr. Michel Reithmueller also of VKI for editing help and Michael Fischer, of AGARD, the director of the project.

In accordance with AGARD format, the acknowledgement for figure contributions is presented here instead of in the figures. I believe that the figures in this type of report are of paramount importance. Sometimes figures such as these evolve over many years and cannot be created by any one individual. I am especially pleased with many of the figures and I took advantage of your ingenuity in creating them. I thank you for these excellent contributions. In most cases the listed individual briefed me on the specific application and provided the figure although the figures do represent contributions of other individuals as well.

Figure 1.18 - Lawrence Livermore/Sandia National Combustion Laboratories - Combustion Workshop Notes.
Figures 2.1-2.4 - USAF AEDC, from the publications of Charles Enis, Paul Dugger, and Roy Hendrix.
Figures 2.5, 2.6 - Spectron, Dr. James Craig.
Figures 2.7, 2.8 - ONERA, from publications of Drs. Claude Veret, Jean Surget, and Michel Philbert.
Figure 2.9 - von Karmen Institute, Dr. Roland Meynart (now at University of Brussels).
Figure 2.10 - Lockheed Marietta Co., Drs. Louis Lorenzo and Clay Whiffen.
Figures 2.22, 2.23 - Aberdeen Proving Ground, Mr. Robert Rowe.
Figures 3.22, 3.23 - NASA Ames Research Center, Drs. Will Nachalo and James Craig.
Figure 3.25 - Wright Aeronautical Laboratories, all from the works of Dr. George Havener (now with S.U.N.Y., Binghamton).
Figure 3.26 - Cambridge University, Dr. Peter Bryanston-Cross.
Figure 3.27 - ONERA, Drs. Claude Veret and Jean Surget.
Figures 3.28-3.30 - Rolls Royce, Drs. Peter Bryanston-Cross, Rick Parker, and David Jones.
Figure 3.32a,b - U.S. Army Aeromechanics Laboratory, Drs. Yung Yu and John Kittleson.

Figure 3.35 - Idaho National Engineering Laboratory, Larry Reynolds and Vance Deason.

Figure 3.32c,d - Spectron, Dr. Hung Tan.

Figures 4.7, 4.26 - French-German Institute at St. Louis (ISL), Drs. G. Smeets and A. George.

Figures 4.19, 4.24, 4.25 - ONERA, Dr. Alaine Boutier and co-workers.

Figure 4.28 - von Karman Institute, Dr. Michel Reichmueller.

Figure 4.29 - University of Nurnberg, Dr. Franz Durat.

Figure 4.30 - NASA Lewis Research Center, Dr. Harold Schock.

Figure 4.31 - ISL, Drs. H. Pheifer and H. Schafer.

Figure 4.32 - ONERA.

Figure 4.33 - Arnold Engineering Development Center, Tennessee, Bobby Nichols.

Figure 4.34 - Spectron, Dr. Cecil Hess.

Figures 4.35, 4.38 - Spectron, Dr. Anthony Smart.

Figures 4.40, 4.41 - Rolls Royce, Dr. David Jones.

Figure 4.42 - ONERA, Dr. Alain Boutier and co-workers.

Figure 5.5 - Stanford University, Dr. Ron Hanson.

Figures 5.8-5.10 - NASA Langley Research Center, Dr. Robert McKenzie.
LASER APPLICATIONS IN FLOW DIAGNOSTICS

J.D. Trolinger
Edited by J.J. Ginoux

October 1988

This is an expansion and update of AGARDograph No. 186, Laser Instrumentation for Flow Field Diagnostics, which was published in 1978. A brief introduction to coherent optics and lasers is given to provide a review and common language to the reader. The material is presented at a level to be understood by engineers who are working in the aerospace field and who are interested in the experimental application of lasers in research and development. An emphasis is placed on applications in aerodynamics, and examples have been selected from a number of laboratories in the NATO countries visited by the author while preparing the material. The applications include holography, laser velocimetry, spectroscopy, and the use of lasers simply as efficient light sources in conventional measurement methods.

This AGARDograph has been produced at the request of the Fluid Dynamics Panel of AGARD.
This is an expansion and update of AGARDograph No.186, Laser Instrumentation for Flow Field Diagnostics, which was published in 1974. A brief introduction to coherent optics and lasers is given to provide a review and common language to the reader. The material is presented at a level to be understood by engineers who are working in the aerospace field and who are interested in the experimental application of lasers in

P.T.O
research and development. An emphasis is placed on applications in aerodynamics, and examples have been selected from a number of laboratories in the NATO countries visited by the author while preparing the material. The applications include holography, laser velocimetry, spectroscopy, and the use of lasers simply as efficient light sources in conventional measurement methods.

This AGARDograph has been produced at the request of the Fluid Dynamics Panel of AGARD.

ISBN 92-835-0480-1