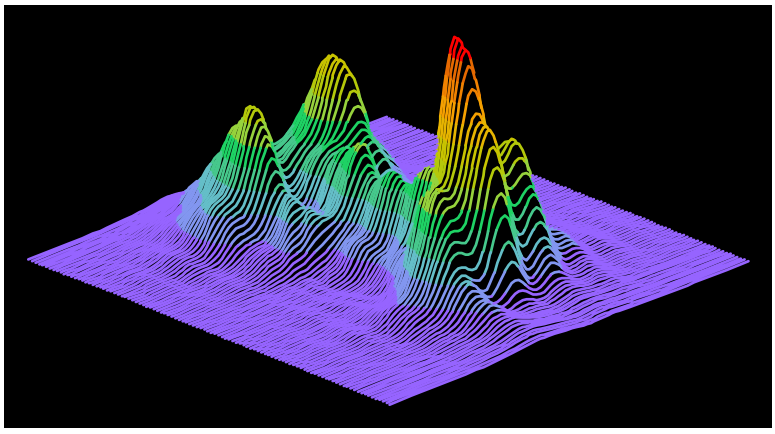
GLAD Theory Manual

Ver. 4.7



Applied Optics Research

Applied Optics Research
1087 Lewis River Rd. #217, Woodland, WA 98674
Tel: (360) 225 9718, Fax: (360) 225 0347, email: glad@aor.com
www.aor.com

Copyright 1986-2001 Applied Optics Research. All Rights Reserved.
Reproduction of this manual is prohibited

Table of Contents

1. Introduction . . . 9
   - 1.1. Documentation . . . 9
   - 1.2. Availability of GLAD . . . 9
2. Background . . . 13
   - 2.1. Review of Physical Optics Modeling . . . 15
     - 2.1.1. Typical Types of Analysis . . . 15
     - 2.1.2. Discussion of Three Dimensional Modeling . . . 16
3. Basic Theory of Propagation . . . 17
   - 3.1. Representation of the Optical Beams . . . 17
   - 3.2. Derivation of the Fundamental Propagation Equations . . . 19
   - 3.3. Diffraction Propagation . . . 20
   - 3.4. The Split Step Method . . . 21
   - 3.5. Angular Spectrum Propagation . . . 24
4. Diffraction Phenomenology . . . 29
   - 4.1. Single Plane Wave Component . . . 33
   - 4.2. Far-Field Diffraction Effects . . . 36
   - 4.3. Gaussian Beams . . . 38
   - 4.4. Significance of Rayleigh distance . . . 39
     - 4.4.1. Finding the Waist . . . 39
     - 4.4.2. Lens Law for Gaussian Beams . . . 39
   - 4.5. Hermite Gaussian Beams . . . 40
5. Propagation in Homogeneous Media . . . 43
   - 5.1. Sampling . . . 47
   - 5.2. Automatic Selection of Matrix Units . . . 50
   - 5.3. Propagation Control . . . 51
     - 5.3.1. Using Surrogate Gaussian Beams for Propagation Control . . . 56
   - 5.4. Finite Difference Propagation . . . 58
     - 5.4.1. Real-Imaginary Description . . . 58
     - 5.4.2. Amplitude-Wavefront Description . . . 59
   - 5.5. Propagation of Rotational Functions and Fast Hankel Transforms . . . 60
6. Optical Systems and Inhomogeneous Media . . . 63
   - 6.1. ABCD Operators . . . 63
     - 6.1.1. Gradient Index Optics . . . 67
     - 6.1.2. System Analysis . . . 69
   - 6.2. Beam Propagation Method . . . 69
   - 6.3. Non-Stigmatic Effects . . . 71
   - 6.4. High Numerical Aperture Lenses . . . 72
   - 6.5. Optical Systems . . . 74

7. Aberration . . . 77
7.1. Aberration Polynomials . . . 77
7.2. Random Aberration . . . 78
7.2.1. Generation of Random Aberration . . . 79
7.2.2. Atmospheric Aberration . . . 80
7.3. Beam Quality Measurement . . . 83
7.3.1. Strehl Ratio . . . 83
7.3.2. Wavefront Variance . . . 83
7.3.3. Optical Transfer Function . . . 84
7.3.4. Autocorrelation Function . . . 84
7.3.5. Irradiance Uniformity . . . 85
7.4. Beam Size Measurement . . . 86
8. Resonators and Lasers . . . 91
8.1. Bare-Cavity Resonators . . . 91
8.1.1. Flat-Flat Resonators . . . 92
8.1.2. Stable Resonators . . . 93
8.1.3. Aligned Unstable Resonator . . . 106
8.1.4. Misaligned Unstable Resonator . . . 108
9. Gain and Nonlinear Media . . . 111
9.1. Beer's Law Gain . . . 111
9.2. Franz-Nodvik, CO_2 Gain . . . 112
9.3. Rate Equation Model . . . 113
9.3.1. Franz-Nodvik Solution . . . 116
9.3.2. Off-line Effects . . . 119
9.3.3. Spontaneous Emission . . . 120
9.4. Three-Level Gain, Ruby Laser . . . 120
9.4.1. Sample Calculation: Three-level Gain, Ruby Laser . . . 123
9.5. Three-Level Gain, Single Upper State . . . 124
9.5.1. Multiple Mode Interactions . . . 127
9.5.2. Sample Calculation: Three-level Gain, Single Upper Level . . . 128
9.6. Raman Modeling . . . 131
9.6.1. Photon-Flux Raman Model . . . 132
9.6.2. Raman Interactions with Multiple Stokes Beams . . . 133
9.6.3. Four-Wave Mixing . . . 133
9.7. Optical Parametric Oscillator . . . 138
9.7.1. Formulation for Efficient Solution . . . 139
9.7.2. Check of Energy Conservation . . . 141
9.7.3. Spatial Frequency Content and Axial Sampling . . . 148
9.8. Sum-Frequency Generation (SFG) . . . 149
9.8.1. Numerical Example, Calculation 1 . . . 152
9.8.2. Numerical Example, Calculation 2 . . . 153
9.9. Birefringent Effects . . . 154
9.9.1. Calculation Steps . . . 156

9.10. Second Harmonic Generation . . . 156
9.11. Transient Raman Kinetics . . . 159
10. Component Specification . . . 169
10.1. Mirrors and Lenses . . . 169
10.2. Tilts and Decenters . . . 170
10.3. Apertures and Obscuration . . . 170
10.4. Complex Systems and GRIN Media . . . 170
10.4.1. Alternate form . . . 171
10.5. Anamorphic ABCD matrices . . . 172
11. Global Coordinate System . . . 175
11.1. Component Location and Rotation in Global Coordinates . . . 175
11.1.1. Code V Translation . . . 178
11.2. Chief Ray Propagation . . . 179
11.2.1. Chief Ray Surface Intercept . . . 179
11.2.2. Calculation of Local Surface Properties . . . 182
11.2.3. Calculation of the Surface Coordinate System . . . 183
11.3. Reflection . . . 184
11.4. Beam Print on Surface . . . 185
11.5. Refraction . . . 186
11.5.1. Exact Ray Tracing . . . 187
11.6. Mirror Systems, Image Rotations, Image Parity, Common Prisms . . . 189
11.6.1. Mirror Matrices . . . 189
11.6.2. Common Mirrors and Prisms . . . 196
11.6.3. Image Positions . . . 196
11.6.4. Beam Print on Tilted Optical Surfaces . . . 201
11.6.5. Accumulating Magnification Over Many Operations . . . 203
11.6.6. Tunnel Diagrams . . . 205
11.6.7. Reduced Distance . . . 205
11.7. Gratings . . . 206
11.7.1. The Bragg Condition . . . 206
11.8. Gratings (Unresolved Microstructure) . . . 208
11.8.1. Phase Grating (Unresolved) . . . 209
11.8.2. Amplitude Grating (Unresolved) . . . 211
11.9. Decomposing Rotation Matrix into Euler Angles . . . 212
12. Atmospheric Propagation . . . 215
12.1. Atmospheric Aberration . . . 215
12.2. Adaptive Optics . . . 219
13. Thermal Blooming . . . 229
13.1. Mathematical Derivations and Techniques . . . 230
13.2. Blooming Algorithms . . . 233
14. Polarization . . . 239
14.1. Rotation of Polarization Components . . . 240
14.1.1. Distributed Effects . . . 241

14.2. Common Polarization Effects and Components . . . 242
14.2.1. Birefringence . . . 242
14.2.2. Reflecting Mirror . . . 242
14.2.3. Optical Activity . . . 243
14.2.4. Faraday Effect . . . 243
14.2.5. Kerr Effect . . . 244
14.2.6. Pockels Effect . . . 244
14.3. Polarization Aberrations . . . 245
14.3.1. Pauli Spin Matrices . . . 246
14.3.2. Polarization Representation in GLAD . . . 248
14.3.3. Polarization Effects of Surfaces . . . 248
14.3.4. Fresnel Reflection with Absorption in the Rarer Medium . . . 256
14.3.5. Goos-Hanchen effect . . . 258
15. Waveguide Grating Couplers . . . 261
15.1. Perturbation Model of Grating Coupling . . . 262
15.2. Induced Dipole . . . 265
15.3. Mode Coupling Theory . . . 270
15.4. TE Guided Mode . . . 272
15.4.1. TM Guided Mode . . . 272
15.4.2. Paraxial Approximation . . . 273
15.5. Bragg Condition . . . 273
15.6. Round Trip: Outcoupling Followed by Incoupling . . . 274
16. Optimization . . . 277
16.1. Method of Least Squares . . . 278
16.1.1. Damping . . . 280
16.1.2. Applying Constraints . . . 280
16.1.3. Lagrange Multipliers . . . 281
16.2. Example of Optimization . . . 283
17. Axicon Mirrors . . . 285
17.1. Radial Mode Propagation . . . 286
17.2. Diffraction Considerations in Generating a Radial Beam by an Axicon . . . 287
17.2.1. Decentration and Tilts . . . 290
17.3. Polarization . . . 292
17.4. Annular Beams . . . 292
17.5. Polarization in Reflaxicon and Waxicon Configurations . . . 297
17.6. Resonators with Axicon Elements . . . 297
17.7. More General Axicon Surfaces . . . 301
18. Partially Coherent Imaging . . . 305
19. Thermal and Stress Effects in Component . . . 309
19.1. Thermal Effects . . . 309
19.2. Thermo-Optical Effects . . . 313
19.3. Thermally Induced Birefringent Stress . . . 315
19.3.1. Governing Equations Of Strain . . . 317

19.3.2. Thermal Stress in a Plane: Two-dimensional Solution 324
19.3.3. Green's Function Method 329
19.4. Optical Effects Of Stress 335
19.4.1. Photoelastic Tensor 340
19.4.2. Simplifying the 3D Calculations 341
20. Integration of Geometrical and Physical Optics 345

Index 349

1. Introduction

GLAD is an advanced physical optics analysis code which may be applied to a wide variety of optical modeling applications. With GLAD, the user is able to model both simple systems and highly complicated, multiple laser configurations. The code is designed to analyze beam trains and laser devices including the effects of diffraction, active media, apertures, lenses and mirrors, and aberration. GLAD has a global coordinate system that allows complex systems to be described and which enables components to be arbitrarily located and rotated.

1.1 Documentation

GLAD is documented in several volumes:

- Introduction to GLAD
- GLAD Theoretical Discussion
- GLAD Command Descriptions
- GLAD Examples

This volume describes the theoretical basis and numerical methods used in GLAD. Detailed explanation of the commands is contained in GLAD Command Descriptions as well as command formats and installation instructions. The volume of examples gives numerous applications of GLAD to illustrate use of the code. Electronic versions of all manuals are on the distribution CD ROM. The most current version of the manuals may be downloaded from www.aor.com.

1.2 Availability of GLAD

GLAD is copyrighted software which is available under license agreement with AOR. Questions regarding GLAD may be addressed to:

Applied Optics Research
1087 Lewis River Rd. #217
Woodland, WA, 98674 USA
tel: 360 225 9718, fax: 360 225 0347
email: glad@aor.com, http://www.aor.com

Publications and Presentations using GLAD

1. Lawrence, G., and P. Wolfe, “Application of the LOTS Computer Code to Laser Fusion Systems,” Los Alamos Scientific Laboratory Conference on Optics'79.
2. Viswanathan, V., I. Liberman, G. Lawrence, and B. Seery, “Optical Analysis of Laser Systems Using Interferometry,” Appl. Opt., 19, 1870 (1978).
3. Wolfe, P., A. Saxman, and G. Lawrence, “LOTS Analysis of Optical Diffraction in Antares,” Los Alamos Scientific Laboratory Conference on Optics'79.

4. Lawrence, G., “Optical Performance Analysis of CO2 Laser Fusion Systems,” doctoral dissertation, University of Arizona, (1980).
5. Lawrence, G., and C. Barnard, “Under sampling Versus Aliasing Errors in Numerical Evaluations Diffraction Patterns,” Los Alamos Conference on Optics '81.
6. Boye, C., W. Swantner, and G. Lawrence, “A Study of Spatial Filtering Using LOTS MCI,” SPIE Conference, August 24, 1982.
7. Barnard, C. and G. Lawrence, “Physical Optics Modeling with Multiple Lasers,” Los Alamos Conference on Optics '83.
8. Lawrence, G., C. Barnard, and V. Viswanathan, “Global Coordinates and Exact Aberration Calculations Applied to Physical Optics Modeling of Complex Optical Systems,” SPIE Proc. Vol. 642 (1986).
9. Knapp C., V. Viswanathan, J. Bender, Q. Appert, G. Lawrence, and C. Barnard, “Analysis of Free Electron Laser Systems with Grazing Incidence Optics,” SPIE Proc. Vol. 642 (1986).
10. Lawrence, G., “Optical System Analysis with Physical Optics Codes,” SPIE Proc. 766–18 (1987).
11. Lange, M. and G. Lawrence, “Numerical developments of diffracting Raman conversions,” SPIE Proc. 739–45 (1987).
12. Lawrence, G. and P. Cronkite, ‘Physical optics analysis of the focusing grating coupler optical data head,’ Fourth Topical Meeting, Optical Data Storage, March 1987.
13. Lawrence, G. and P. Cronkite, ‘Physical optics analysis of the focusing grating coupler optical data head,’ Appl. Opt., 27, 672–678 (1988).
14. Scholl, M., A. Khatib, G. N. Lawrence, “Optical Modeling of a Space Relay Experiment,” OE Lase '88.
15. Lawrence, G., M. S. Scholl, A. Khatib, “Modeling of Dynamic Effects in Propagation of a Low Power Laser Beam,” OE Lase '88.
16. Lawrence, G., M. Wang, S H Hwang, “Optical Performance of a Waveguide Grating Imaging Device,” SPIE Proc. 965–06 (1988).
17. Lawrence G., B. Herman, and A. Ben David, “Effects of Atmospheric Scattering on Irradiance Distribution of a Downward Propagating Laser Beam,” SPIE Pro. 965–15 (1988).
18. Eckhardt, S., G. Lawrence, J. Burke, “Diffractive and Aberration Effects of Propagation through a Refractive Surface,” SPIE Proc. 1045–28 (1989).
19. Lawrence, G., “Polarization Modeling in Physical Optics Analysis,” SPIE 33rd Annual Symposium, August 1989, 1166–08.
20. Hwang, S H and G. Lawrence, “Physical Optics Analysis of Gradient Index Optics,” SPIE 33rd Annual Symposium, August 1989, 1168–40.
21. Daniel R. Neal, William C. Sweatt and George N. Lawrence, “Software models complex laser systems,” Laser Focus World, p59, June 1990.
22. G. Lawrence, “Physical Optics in Optical Design,” SPIE Proc. Vol. 1354 (1990).
23. G. Lawrence, “Advances in optics software,” Optics and Phtonics News, 31–34, Sep. (1990).

24. G. Lawrence and K. Moore, “Integrating Geometrical and Physical Optics,” SPIE Proc. 1415-33 (1991).

25. M. S. Scholl and G. N. Lawrence, “Diffraction modeling of a space relay experiment,” Opt. Eng. 39 (3), 271–278 (1990).

26. Y. Lin, T. Kessler, G. Lawrence, “Raman scattering in air: A four dimensional system analysis,” SPIE Proc. Vol. 1625.

27. G. Lawrence, “Computer aided Optimization of Laser Systems,” SPIE Proc. Vol. 1625.

28. G. Lawrence, “How to select lens design software,” Laser Focus World, 111–116, July, 1993.

29. Y. Lin, T. J. Kesler, J. J. Armstrong, and G. Lawrence, “Laser system power balance effects from stimulated rotational Raman scattering in air,” SPIE Proc. Vol. 1870, 14–25 (1993).

30. G. Lawrence, “Using Rules of Thumb in the Design of Physical Optics Systems,” Conference on Optical Photonics, Palm Springs, March, 1993.

31. Weng. W. Chow, William. C. Sweatt, George N. Lawrence, Tanya E. Jewell, “Physical Optics Modeling of Soft X ray Projection Lithography (SXPL) Imaging Experiments,” OSA Technical Digest, Conf. on Soft X ray Projection Lithography, May 10 22, Monterrey, 1993.

32. G. Lawrence and T. Baer, “Characterization of Laser Beam Quality by M^2”, SPIE Proc. Vol. 2117, Jan. 1994.

33. G. Lawrence, “Proposed international standard for laser beam quality falls short,” Laser Focus World, p. 109–114, July, 1994.

34. M. S. Scholl, G. N. Lawrence, “Adaptive optics for in orbit aberration correction — feasibility,” Appl. Opt., 34 , 7295–7301 (1995).

35. Y. Lin, T. J. Kessler, and G. N. Lawrence, “Design of Continuous Surface relief Phase Plates by Surface based Simulated Annealing to Achieve Control of Focal Plane Irradiance,” Opt. Lett. 21, 1703 –1705 (1996).

36. Y. Lin, T. J. Kessler and G. N. Lawrence, “Distributed Phase Plate for Supergaussian Focal Plane Irradiance Profiles with Low Scattering Loss,” Opt. Lett. 20, 764–766 (1995).

37. Y. Lin, T. J. Kessler, and G. N. Lawrence, “Raman Scattering in Air: A Four Dimensional Analysis,” Appl. Opt. 33, 4781–4791 (1994).

38. G. Lawrence, “When Fourier methods fail — diffraction propagation through thick elements and tilted surfaces,” Presented to Workshop on Asymptotic and Approximate Methods in Optical Modeling, The Rochester Theory Center for Optical Science and Engineering, Rochester, NY June 12–13, 1997.

39. Ying Lin and Jesse Buck, “Numerical modeling of the excimer beam,” SPIE Vol. 3677, 700–710 (1999).

40. G. Lawrence, “Integrating geometrical and physical optics with the lensgroup operator method,” SPIE Proc. 3780–06 (2000).

2. Background

As the complexity and variety of laser systems has expanded over the years, the need for powerful analytical methods has become increasingly important. Optical engineers and scientists need to be able to accurately calculate performance in order to optimize designs and to determine system tolerances. Numerical analysis is quicker and less expensive than laboratory experiments and also serves as an educational tool. The optical engineer or scientist can determine the end to end performance of a complex device based on the characteristics of the lenses and mirrors, propagation distances, apertures, aberrations, laser gain, and other effects. The definition of the system components can be very detailed, including exact aperture shapes and the accurate aberration determined from interferometry or other means of measurement. A relatively complete description of the laser beam can be determined by the intensity and phase profiles. This information can be used to find the total power, the peak power, wavefront quality, wavefront variance, Strehl ratio, and properties of the focused beam.

Originally, optical modeling was used to understand relatively simple laser systems and beam trains. More recently much more complex systems have been analyzed which include atmospheric aberration, adaptive optics, phased arrays, nonlinear optics, etc. Optical modeling is a highly complex field and a single article can only begin to touch on the important areas of the subject. The most critical and difficult aspect of optical modeling is the general treatment of diffraction propagation in homogeneous media and in systems of lenses, mirrors, and inhomogeneous media. A complete and consistent mathematical description and set of algorithms are presented in this article that can treat all types of systems to the accuracy of small angle, scalar Fresnel diffraction theory. Split-step methods for treating nonlinear gain and aberration are described. Optical modeling is a relatively new field of optics going back to the earliest days of the development of the laser. Unlike geometrical optics which had a noble history before the development of the computer, optical modeling, with its intensive calculations and large memory requirements is a child of the computer age. Computers operating in the megaflop range (millions of floating point operations) are now readily affordable and commonly available and there seems to be no diminution of the rate of improvement of the computer hardware. A few years ago, only major organizations could afford the computer equipment and the staff of experts to develop and operate physical optics codes. The development of user oriented programs, has made it possible for the optical scientist or engineer to successfully perform complex calculations without being a code specialist.

Prior to the advent of lasers, optical analysis consisted largely of geometrical ray tracing for the design of photographic systems. Diffraction analysis was applied to various types of apertures from the turn of the century but generally not to the analysis of systems. The laser was the stimulus for physical optics calculations. In geometrical analysis, the light is represented by a set of rays which are normal to the wavefront as shown in Fig 2.1. For short propagation lengths such as are encountered in a common photographic lens, the diffraction effects are small and localized to the edge of the beam. For this type of problem, rays do a good job of determining the aberrations of the system and a reasonably good job of determining the intensity variations. For a conventional optical system the rays enable us to calculate the aberrations. These aberrations may be used to determine the pupil function and a simple far field diffraction analysis may be made or, if the system is not diffraction limited, the rays may be traced to the image plane

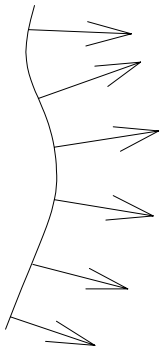


Fig. 2.1. Representation of an optical beam by rays.The rays convey optical path differences errors and slight differences in ray direction indicate ray aberrations.

and the geometric image size may be used as a measure of performance. Associating equal energy with each ray, we can get a rough estimate of the energy density but this method breaks down the region of the focus.

A simple spatial filter is commonly used in many laser systems. Ray optics can approximately calculate the image at the focal point where the pinhole filter is placed. Ray optics is unable to predict removal of the phase aberrations by the spatial filtering and smoothing of the intensity distribution. Figure 2.2 shows two lenses and a pinhole aperture at the intermediate focus which acts as a spatial filter. This simple device is used in many laser systems to remove the aberrations and to smooth out intensity variations. Using geometrical analysis, we may be able to approximately determine the image size but we cannot determine the reduction of aberrations and the change in the intensity distribution to be found in the expanded beam after the spatial filter. The spatial filter can not be analyzed by geometrical analysis. Physical optics analysis describes the optical beam by a complex amplitude function, describing the transverse beam distribution. The complex amplitude includes both the intensity and phase information of the beam at one axial position. This information can be modeled in the computer by a complex two dimensional array where each point of the array corresponds to a point on the beam.

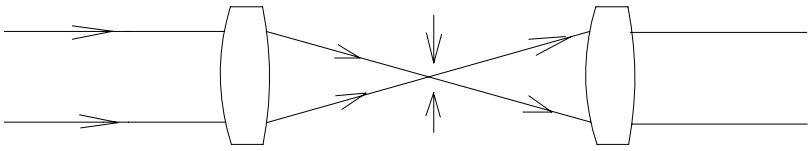


Fig. 2.2. A simple spatial filter is commonly used in many laser systems. Rays optics can approximately calculate the image at the focal point where the pinhole filter is placed. Ray optic is unable to predict removal of the phase aberrations by the spatial filter and smoothing of the intensity distribution.

The earliest work in resonator analysis codes was done for optical communications in the 1960's by Fox and Li[1]. The military interest in high energy lasers stimulated intense development of physical optics modeling codes in the mid 1970's. The work by Siegman and Sziklas in 1974 and 1975 studied gas dynamic lasers including diffraction, the active gain medium, apertures, and aberration. The first paper by Siegman

and Sziklas used an Hermite Gaussian expansion for propagation [2]. The second paper by Sziklas and Siegman used a fast Fourier transform (FFT) method for propagation [3]. A third method based on finite difference propagation — a direct solution to the differential equation of diffraction — was used by Rench and Chester in 1974 [4]. Over time, the FFT method has become the mainstay of optical propagation codes, as much for its modest and well understood sensitivity to error as for its computational efficiency for many types of problems.

2.1 Review of Physical Optics Modeling

GLAD is designed to calculate the performance of laser systems and other optical systems which have a well defined direction of propagation. GLAD represents the optical beam by the complex amplitude of the optical wavefront. This is distinct from geometrical optics codes which represent the optical beam by rays. Geometrical codes are very useful when near field diffraction and gain are not important and where the transverse intensity distribution of the beam is either constant or some simple function.

GLAD is designed to treat entire systems from end to end. A full end-to-end system analysis code must necessarily simplify to some extent the components and physics to keep development costs reasonable and to achieve satisfactory run times. In many cases, specialty codes are developed to address in greater detail components or processes of particular interest. Knowledge gained from specialty codes may be used to form simpler, more efficient models for use in end to end modeling.

2.1.1 Typical Types of Analysis

To illustrate application of the code, consider the schematic shown in Figure 2.3. The configuration does not represent any particular system, although it has some resemblance to a Raman amplifier. Many of the important features of laser systems are present. Beam 1 (red) is shown starting with a quasi-gaussian distribution. GLAD allows several different ways of defining the starting distribution. GLAD assumes no particular symmetry to the optical distribution. Also, the distribution may be decentered.

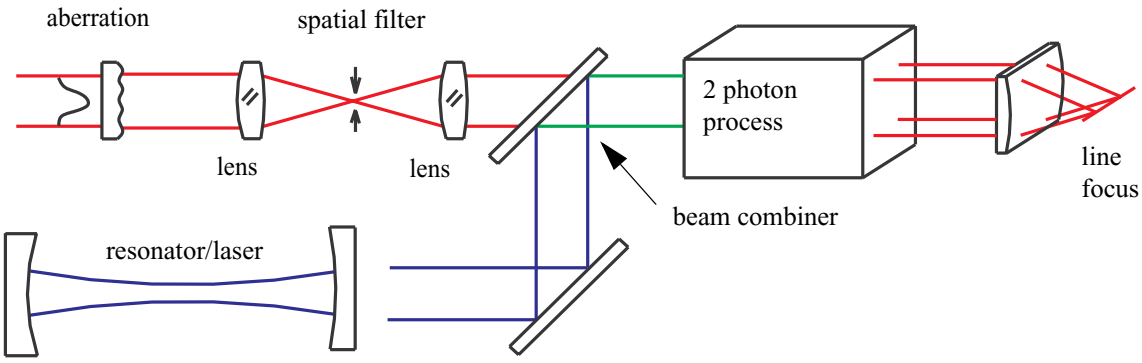


Fig. 2.3. Schematic for Raman amplification.

Beam 1 is shown with an aberrated element. A large variety of types of aberration may be used in GLAD. The lens in Beam 1 brings the light to a focus. GLAD may be used to calculate the distribution at any point in the collimated, converging, or diverging part of the beam. An aperture at the focus of the lens

acts to spatially filter the distribution to remove some of the aberration. The second lens recollimates the beam.

Beam 2 (blue) is being generated by a laser. GLAD can model most types of stable and unstable resonators and a variety of gain media. Beam 2 is added to Beam 1 by a beam combiner. The combination may be done coherently or incoherently. Coherent addition is possible only for beams having nearly identical wavelengths. The combined beams interact in a second medium. This might be a Raman scattering cell.

After the two photon process, a cylindrical lens is shown. Glad may be used to model spherical, cylindrical, or toroidal optical elements. GLAD is designed to be very modular. With basic modules such as diffraction steps, lenses, mirrors, apertures, beam splitters, beam combiners, and gain media; a large variety of optical systems may be analyzed.

2.1.2 Discussion of Three Dimensional Modeling

GLAD was developed as a three dimensional code - modeling two transverse dimensions by the two dimensional computer arrays and the axial dimension by successive calculations. In general, a four dimensional solution may be required because of temporal dependence of the optical beam. Many systems may be approximated by steady state solution, because the temporal pulse width is longer than the time constants of all processes in the system. GLAD is well suited to analyze this type of problem. Other systems have short temporal pulse widths. If the pulse width is shorter than the time constants of all the processes in the system, then the exact waveform of the pulse does not play a role: only the integrated effect need be used in modeling. The beam may be represented by fluence in terms of joules per square centimeter.

With the increasing availability of very fast workstations, time dependent, three dimensional problems may be solved in reasonable time. Example 27 illustrates examples of jitter including the pulse to pulse variation and time integration of the energy. Example 56 illustrates the spectral performance of a Fabry Perot cavity.

The most difficult problems are ones where the temporal pulse shape plays an important role. GLAD has been applied to a variety of transient, three dimensional problems. Example 79 illustrates transient Raman analysis and Ex. 80 describes a time dependent analysis of a Q switched laser. Partial coherent effects are shown in Ex. 83.

References

1. A. G. Fox and T. Li, “Resonant Modes in a Maser Interferometer,” Bell System Technical Journal, Vol. 46, 453 (1961).
2. A. E. Siegman, “Hermite Gaussian Functions of Complex Argument as Optical Beam Eigenfunctions,” JOSA Vol. 63, 1093 (1973).
3. E. A. Sziklas and A. E. Siegman, “Diffraction calculations using fast Fourier transform methods,” Proc. IEEE 62, 410–412 (1974).
4. D. B. Rench and Chester, “Three dimensional unstable resonators with laser medium,” Appl. Opt. 13, 2546–2561 (1974).

3. Basic Theory of Propagation

GLAD is designed to allow the user to perform optical calculation by entering the optical configuration information and by defining the starting optical beam parameters. The code then performs the optical transformations to simulate the various components and phenomena in the device. The code can be operated with only a rudimentary understanding of propagation theory. However, as with any other scientific or engineering field, a good understanding of the analytical tools leads to greater insight into the problems being studied.

3.1 Representation of the Optical Beams

The optical beam may be represented as a three dimensional time varying vector field,

$$\boldsymbol{E}(x, y, z, t) = E_x(x, y, z, t)\hat{\boldsymbol{i}} + E_y(x, y, z, t)\hat{\boldsymbol{j}} + E_z(x, y, z, t)\hat{\boldsymbol{k}}, \tag{3.1}$$

where boldface indicates vector quantities.

Optical beams with convergence cones of f number greater than about 1.5 are well described without the E_z term. GLAD uses only the E_x and E_y terms. The E_x and E_y terms represent orthogonal states of polarization, with separate computer arrays for each polarization state as shown in Fig. 3.1. By defining the relative amplitudes and phase differences between E_x and E_y, various states of polarization can be defined: linear, circular, and general elliptical polarizations. Where different polarization states are not required, GLAD performs calculations only on E_x.

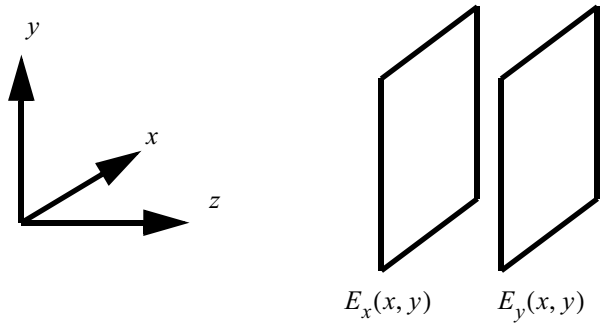


Fig. 3.1. Orthogonal polarization fields, E_x and E_y.

GLAD defines two dimensional computer arrays that represent the transverse distribution of the optical beam at a specific point. The representation is therefore

$$\boldsymbol{E}(x, y, z, t) = E_x(x, y, z, t)\hat{\boldsymbol{i}} + E_y(x, y, z, t)\hat{\boldsymbol{j}}. \tag{3.2}$$

The semicolons indicate variables which are not incorporated into the computer storage arrays, but which are retained as separate parameters in the code. The dependency of the optical beam on time may be neglected in many cases; either because the optical beam is so slowly varying that only the steady state solution is needed or because the pulse is so short that all physical processes in the system see only the integrated effects of the optical pulse. In the steady state solution, temporal dependence can be removed and only the z dependence retained. A single computer array is sufficient (two arrays if polarization effects are included). In the case where the pulse is very short, the E(x,y;z;t) term may be replaced by an integrated amplitude term such that

$$|\boldsymbol{E}(x, y, z)|^2 = \int |E(x, y, z, t)|^2 dt \tag{3.3}$$

and the phase of $E(x, y, z)$ represents the optical path associated with the $E(x, y, z)$ point. The irradiance, or fluence values for a pulsed system, are defined to be

$$I(x, y, z) = E_x(x, y, z)E_x^*(x, y, z) + E_y(x, y, z)E_y^*(x, y, z), \tag{3.4}$$

where $E^*(x, y)$ is the complex conjugate of $E(x, y)$, and $I(x, y)$ is the irradiance. Note that E_x and E_y are amplitudes of orthogonal polarization states.

Where the temporal variation of the optical field must be taken into consideration, the immediate question is whether the pulse shape is constant over the transverse dimensions of the field. If so, the problems are separable and the pulse shape may be moved into kinetics routines. If not, then multiple arrays must be defined, one for each temporal section to be studied. The representation in discrete complex amplitude arrays is

$$E(x, y, z, t) = \sum_i E(x, y, z, t_o + i\Delta t). \tag{3.5}$$

A uniform separation in time of Δt is assumed.

Alternately, the expression may be expanded in frequency with each array representing a distinct wavelength

$$E(x, y, z, \omega) = \sum_i E(x, y, z, \omega_o + i\Delta\omega). \tag{3.6}$$

GLAD allows up to 128 optical beams to be used simultaneously. Since computational burden is directly proportional to the number of beams, the fewest possible should be used. In particular, multiple longitudinal modes will generally have the same transverse structure. The spectral bandwidth effects may be approximated by adroit choice of kinetics parameters.

Problems such as stimulated Raman scattering, stimulated Brillouin scattering, second harmonic generation, and nonlinear optical phase conjugation require separate arrays for the constituent beams.

3.2 Derivation of the Fundamental Propagation Equations

The evolution of the optical fields is a function of diffraction and the gain and loss mechanisms in the beam train. For the sake of completeness, we begin with Maxwell's equations as they apply to a laser medium or free space propagation. We then show the simplifying assumptions needed to get the inhomogeneous wave equation which will be solved by numerical techniques. For a thorough treatment of the subject the reader is referred to one of the many excellent texts on laser physics such as Sargent, Scully, and Lamb [1].

The equations derived pertain to a laser field in a weakly perturbing medium. In MKS units the equations are

$$\nabla\times\boldsymbol{E} = -\frac{\partial\boldsymbol{B}}{\partial t}, \tag{3.7}$$

$$\nabla\times\boldsymbol{H} = \boldsymbol{J}+\frac{\partial\boldsymbol{D}}{\partial t}, \tag{3.8}$$

$$\nabla\cdot\boldsymbol{D} = \rho, \tag{3.9}$$

$$\nabla\cdot\boldsymbol{B} = 0, \tag{3.10}$$

$$\boldsymbol{D} = \varepsilon_0\boldsymbol{E}+\boldsymbol{P}, \tag{3.11}$$

$$\boldsymbol{B} = \mu\boldsymbol{H}, \tag{3.12}$$

$$\boldsymbol{J} = \sigma\boldsymbol{E}, \tag{3.13}$$

where $\boldsymbol{E}$ is the electric field, $\boldsymbol{B}$ is the magnetic induction, $\boldsymbol{D}$ is the electric displacement, $\boldsymbol{H}$ is the magnetic field intensity, ε_0 is the electric permittivity of free space, μ is the magnetic permeability of the medium, σ is the conductivity of the medium, and ρ is the charge density.

Taking the time derivative of Eq. 3.8 and substituting it into the curl of Eq. 3.7, we have

$$\nabla\times\nabla\times\boldsymbol{E} = \mu\sigma\frac{\partial\boldsymbol{E}}{\partial t}-\mu\varepsilon_0\frac{\partial^2\boldsymbol{E}}{\partial t^2}-\mu\frac{\partial^2\boldsymbol{P}}{\partial t^2}, \tag{3.14}$$

$$\nabla\times\nabla\times\boldsymbol{E} = \nabla(\nabla\cdot\boldsymbol{E})-(\nabla\cdot\nabla)\boldsymbol{E}. \tag{3.15}$$

$$\nabla(\nabla\cdot\boldsymbol{E}) = \frac{1}{\varepsilon_0}\nabla(\nabla\cdot\boldsymbol{D}-(\nabla\cdot\boldsymbol{P}))\approx 0 \tag{3.16}$$

In Eq. 3.16, it has been assumed that $\nabla\cdot\boldsymbol{D}$ and $\nabla\cdot\boldsymbol{P}$ are negligible. This is a good assumption for many laser media. Equation 3.15 can be simplified using Eq. 3.16 to under the condition $\sigma\approx 0$,

$$\nabla^2 \boldsymbol{E} = \mu\varepsilon_0 \frac{\partial^2 \boldsymbol{E}}{\partial t^2} + \mu \frac{\partial^2 \boldsymbol{P}}{\partial t^2}. \tag{3.17}$$

The beams to be considered in GLAD will have a well defined direction of propagation along some axis, which we may define as the z axis. Diverging or converging beams may be treated as collimated beams with the base radius of curvature and scaling properties treated separately, as will be discussed below.

In Eq. 3.17, both $\boldsymbol{E}$ and $\boldsymbol{P}$ are rapidly varying in time and along the z axis. Our primary interest is in the slowly varying envelopes for both quantities. This can be done by representing the two vectors as summations of the frequencies composing the optical beam.

A specific frequency component has the form

$$\boldsymbol{E}(x, y, z, t) = \sum_i \boldsymbol{E}_i(x, y, z) e^{-j(\omega_i t - \boldsymbol{k}_i \cdot \boldsymbol{z})}, \tag{3.18}$$

where $k_i = \omega_i(\mu\varepsilon)^{1/2} = 2\pi/\lambda_i$ and $\varepsilon = \varepsilon_0 n^2$.

The term $\boldsymbol{E}_i$ is slowly varying in time. Similarly, the polarization term may be separated into its slowly varying part and a rapidly varying exponential. The polarization term commonly takes the form

$$\boldsymbol{P} = \varepsilon_0 \chi \boldsymbol{E}. \tag{3.19}$$

For nonlinear optical effects, the medium polarization takes the form

$$P = \varepsilon_0 \chi_{ijk} \boldsymbol{E}_i \boldsymbol{E}_j \boldsymbol{E}_k^* \exp[j((\boldsymbol{k}_i + \boldsymbol{k}_j - \boldsymbol{k}_k) \cdot \boldsymbol{z}) - j(\omega_i + \omega_j - \omega_k)t]. \tag{3.20}$$

For all cases of practical interest, $\omega_j \approx \omega_k$ and $k_j \approx k_k$. The slowly varying polarization term, $\boldsymbol{P}_i$, can be used for each temporal frequency component of $\boldsymbol{E}$,

$$\boldsymbol{P}(x, y, z, t) = \sum_i \boldsymbol{P}_i(x, y, z, t) e^{j(k_i z - \omega_i t)}. \tag{3.21}$$

3.3 Diffraction Propagation

For three dimensional calculations, the representation of Eq. (3.2) is used. It is more traditional to use E for the electric field for E&M calculations but for the remainder of this discussion lower case "a" will be used for the complex amplitude, which is related to irradiance by $\boldsymbol{I} = |a|^2 = (nc\varepsilon_0/2)|E|^2$, where c is the speed of light in vacuum and $n = \sqrt{\varepsilon/\varepsilon_0}$. The most common codes use scalar Fresnel diffraction theory where the a_x and a_y terms represent orthogonal states of polarization, with separate computer arrays for each polarization state. The representation is therefore

$$\boldsymbol{A}(x, y; z, t) = a_x(x, y; z, t)\hat{\boldsymbol{i}} + a_y(x, y; z, t)\hat{\boldsymbol{j}}, \tag{3.22}$$

where the semicolons indicate variables which are not incorporated into the computer storage arrays, but which are retained as separate parameters in the code. By defining the relative amplitudes and phase differences between a_x and a_y, various states of polarization can be defined: linear, circular, and general elliptical polarizations. Where different polarization states are not required, calculations may be performed using only on one array. Most physical optics code defines two dimensional computer arrays that represent the transverse distribution of the optical beam at a specific axial point.

The majority of work has been in development of time dependent, one dimensional codes for detailed analysis of nonlinear optical effects or steady state, three dimensional codes for analysis of complex optical systems. Full time dependent, three dimensional models (true four dimensional models) are practical on machines of the 10 megaflop to 100 megaflop machines but are only in the early stages of development. Development of full four dimensional codes must draw upon the time independent, three dimensional methods, which are the subject of this article.

The dependency of the optical beam on time may be neglected in many cases; either because the optical beam is so slowly varying that only the steady state solution is needed or because the pulse is so short that all physical processes in the system see only the integrated effects of the optical pulse. In either case, we can drop the time dependence,

$$\boldsymbol{A}(x, y; z) = a_x(x, y; z)\hat{\boldsymbol{i}} + a_y(x, y; z)\hat{\boldsymbol{j}}. \tag{3.23}$$

Many problems require multiple optical beams to be used simultaneously. Since computational burden is directly proportional to the number of beams, the fewest possible should be used. Problems such as stimulated Raman scattering, stimulated Brillouin scattering, second harmonic generation, and nonlinear optical phase conjugation require separate arrays for the constituent beams.

3.4 The Split Step Method

The evolution of the optical fields is a function of diffraction and the gain and loss mechanisms in the beam train. For a detailed derivation from Maxwell's equations the reader is referred to one of the many excellent texts on laser physics such as Sargent, Scully, and Lamb[1]. Among the simplifications are the rotating wave approximation and the assumption that second derivative of $\boldsymbol{A}$ is slowly varying along the axis, i.e.,

$$\left|\frac{\partial^2 \boldsymbol{A}}{\partial z^2}\right| \ll \left|k\frac{\partial \boldsymbol{A}}{\partial z}\right|. \tag{3.24}$$

For many lasers, the differential equation for the optical field may be written,

$$\frac{\partial \boldsymbol{A}}{\partial z} = -j\frac{1}{2k}\nabla_\perp^2 \boldsymbol{A} - j\frac{\mu\omega^2}{2k}\boldsymbol{P}, \tag{3.25}$$

ignoring the time variation of $\boldsymbol{A}$ and dropping $e^{-j\omega t}$. For nonlinear optical effects, the medium polarization may take a more complex form, according to Bloembergen[2],

$$\boldsymbol{P} \propto \chi \boldsymbol{A} + \chi^{(2)} \boldsymbol{A}\boldsymbol{A} + \chi^{(3)} \boldsymbol{A}\boldsymbol{A}\boldsymbol{A} + \ldots , \tag{3.26}$$

where the superscripts indicate the linear and various higher order nonlinear susceptibilities. For example the polarization term for four wave mixing takes the form,

$$\boldsymbol{P} \propto \chi_{ijk} \boldsymbol{A}_i \boldsymbol{A}_j \boldsymbol{A}_k^* e^{j(\boldsymbol{k}_i + \boldsymbol{k}_j - \boldsymbol{k}_k \cdot \boldsymbol{z})} . \tag{3.27}$$

For linear media, Eq. (3.19) then takes the form,

$$\frac{\partial \boldsymbol{A}}{\partial z} = -j\frac{1}{2k}\nabla_\perp^2 \boldsymbol{A} - j\frac{k\chi}{2n^2}\boldsymbol{A} , \tag{3.28}$$

where n is the index of refraction in the medium. Equation (3.28) describes the propagation of a laser beam in gain media. The first term on the right is the diffraction term; the second is the effect of the medium. In general this equation can not be solved in closed form. Numerical methods are well understood for solving each of the terms on the right if taken separately (Hardin and Tappert[3]).

The conclusion to be drawn from this derivation is that for small steps we can separate the effects of diffraction and kinetics. We can develop our theory for these two areas completely separately and we can perform the calculation in serial fashion where we alternate the two types of calculation. It is the nonlinear aspect of the gain that requires the iterative treatment. Linear gain merely scales the distribution.

In applying a finite difference solution to Eq. (3.28), the field after a small propagation z is given by

$$A(z + \Delta z) = A(z) + \Delta A . \tag{3.29}$$

For small steps, the term A may be separated into a diffraction, and a medium term,

$$\Delta A = \Delta A_{diff} + \Delta A_{medium} , \tag{3.30}$$

$$\Delta A_{diff} = -j\frac{1}{2k}\nabla_\perp^2 A \Delta z \text{ and } \Delta A_{medium} = -j\frac{k\chi}{2n^2} A \Delta z . \tag{3.31}$$

In practice the preferred method of solution of Eq. (3.31) is by FFT, but the concept of taking short steps to limit the change due to diffraction still holds. The solution of Eq. (3.31) is performed in separate steps as shown in Fig. 3.2.

Even when the medium has nonlinear gain or absorption, the effect of the nonlinearities on diffraction effects is often relatively modest. We can reduce the errors to an acceptable level by taking short steps through the medium. This method is often referred to as the split step method. Figure 3.3 illustrates schematically how the same distance may be traversed by different split step choices.

The results of the calculations are shown in Figures 3.4a–3.4d: diffraction without gain (Fig. 3.4a), one step (Fig. 3.4b), two steps (Fig. 3.4c), and four steps (Fig. 3.4d). This illustrates qualitatively that the errors of inadequate axial sampling are often not too severe. It is, therefore, a workable procedure to break the calculation into a modest number of steps and to cover the distance twice: once considering only diffraction

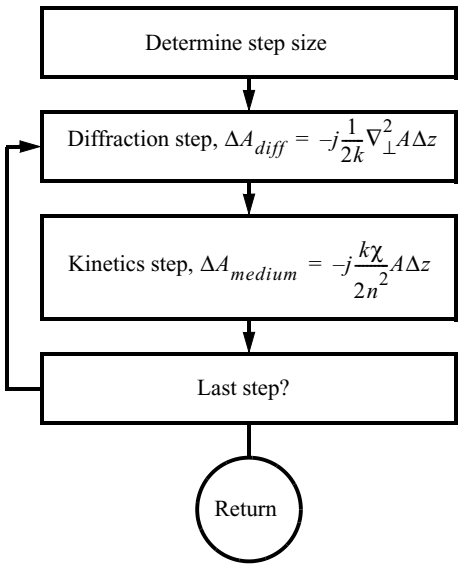


Fig. 3.2. Flow chart for diffraction and kinetics routines. Diffraction calculations assume no gain and may be done by FFT's or finite-difference calculations. Kinetics calculations assume no diffraction. The split-step procedure gives a good approximation for a small step and may be repeated to accomplish arbitrarily long distances.

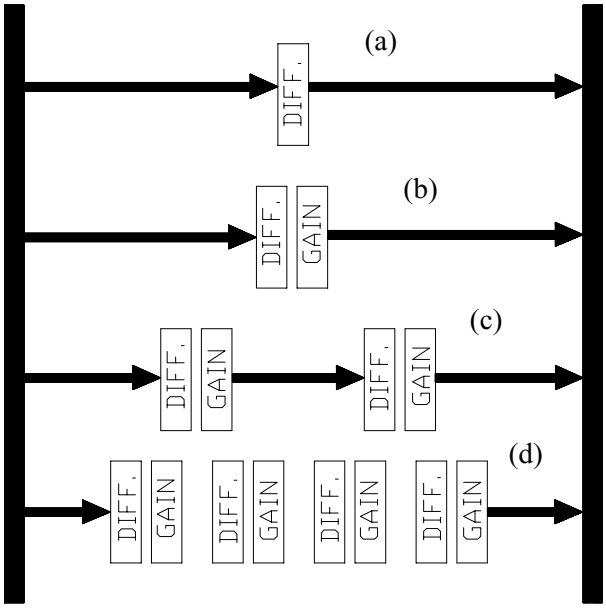


Fig. 3.3. Split-step methods shown to cover the same distance for the four cases shown in a-d. (a) shows diffraction without gain, (b) shows one split-step of diffraction and gain; (c) shows two split-steps; and (d) shows four split-steps. (a)-(d) correspond to Figs. 3.4a–3.4d.

and the second time considering only gain. We can always increase the number of steps and see if the calculations change. Later we shall consider somewhat more quantitative means of determining the number of steps required.

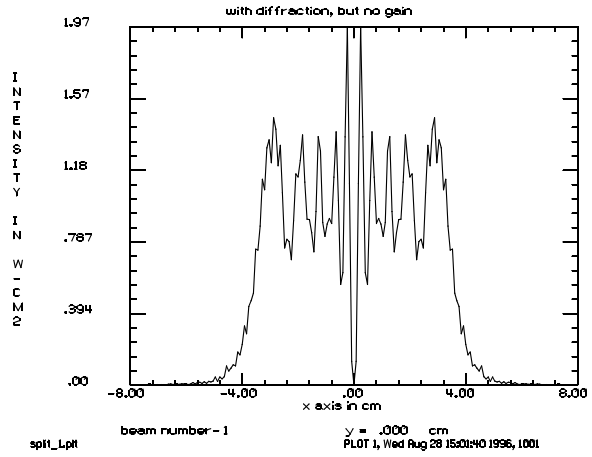


Fig. 3.4a. Diffraction propagation of a beam with only linear gain.

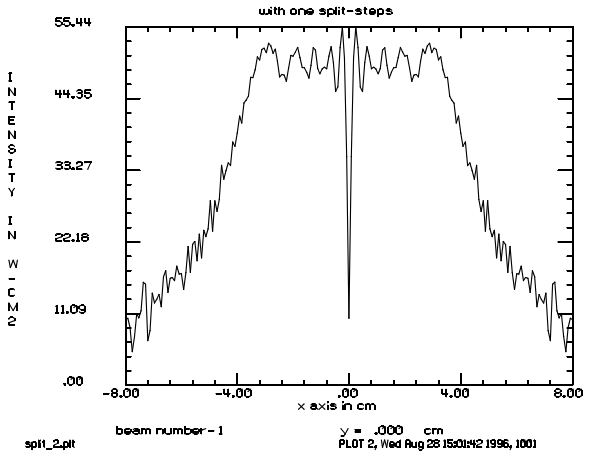


Fig. 3.4b. Diffraction and nonlinear gain (one step). The diffraction ripples are strongly saturated.

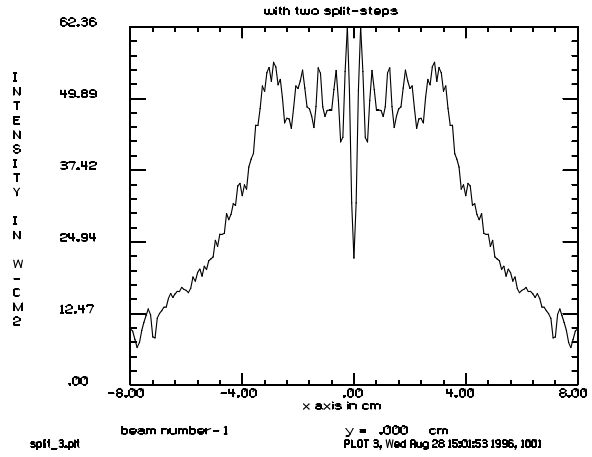


Fig. 3.4c. Diffraction and nonlinear gain (two steps). The saturation is much less than putting all the saturation at the end.

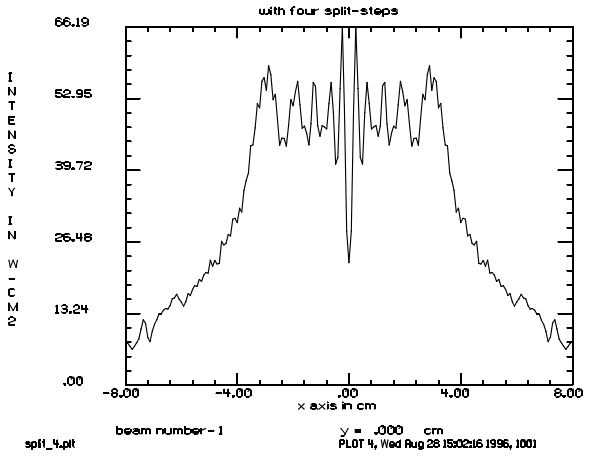


Fig. 3.4d. Diffraction and nonlinear gain (four steps). Using four steps is almost the same as two steps.

3.5 Angular Spectrum Propagation

In homogeneous media, plane waves retain their identity as plane waves during propagation. Plane waves are, therefore, eigenfunctions of diffraction propagation. The eigenvalues are $\exp(j\boldsymbol{k}\cdot\boldsymbol{r})$, where $\boldsymbol{k}$ is the vector wavenumber and $\boldsymbol{r}$ is the reference direction, commonly the z axis. The magnitude of $\boldsymbol{k}$ is $2\pi/\lambda$, where λ is the wavelength in the medium, $\lambda = \lambda_0/n$, where λ_0 is the wavelength in vacuum. The equations will be developed in terms of the wavelength in the medium.

One very effective method of calculating diffraction propagation of an arbitrary complex amplitude distribution is to decompose the distribution into a summation of plane waves, propagate the plane waves

individually using the eigenvalues, and resume the plane waves. This procedure is called the angular spectrum decomposition method [4].

We can compare the geometrical representation of the wavefront and propagation with the complex amplitude and angular spectrum propagation. Geometrical rays are normals to the wavefront. We use enough rays to sample the wavefront thoroughly. We might for example trace an optical system using hundreds of rays. The ray direction is defined by wave number unit vector, $\hat{\boldsymbol{k}}$, with direction perpendicular to the wavefront. For free space propagation of along the ray a distance q, the ray position vector is transformed,

$$\boldsymbol{r}_2 = \boldsymbol{r}_1 + q\hat{\boldsymbol{k}}. \tag{3.32}$$

Propagation of a plane wave is very similar to geometric propagation. A plane wave of amplitude $A(\boldsymbol{k})$ is propagated by the equation,

$$A(\boldsymbol{k}; z) = A(\boldsymbol{k}; 0)e^{j\boldsymbol{k}\cdot z}. \tag{3.33}$$

Geometrical and physical optics propagations are illustrated schematically in Figs. 3.5a and 3.5b. The propagation distance depends on the direction of the plane wave. Evaluating the phase along the z axis, we have

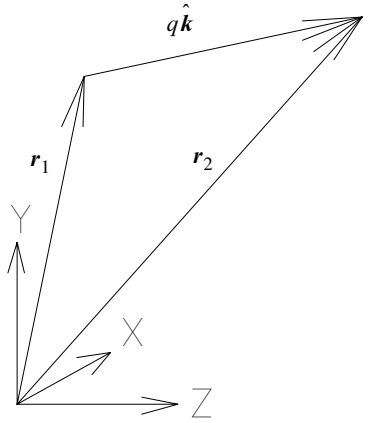


Fig. 3.5a. Geometrical propagation.

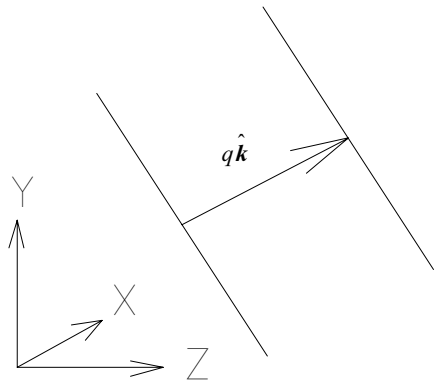


Fig. 3.5b. Physical optics propagation.

$$e^{j\boldsymbol{k}\cdot\boldsymbol{z}} = e^{jk_z z}, \tag{3.34}$$

where k_x, k_y, and k_z are the components of the wavenumber vector.

$$k_x^2 + k_y^2 + k_z^2 = |\boldsymbol{k}|^2 = k^2. \tag{3.35}$$

We can make the approximation (Goodman[4])

$$e^{jk_z z} = e^{jz\sqrt{k^2(1-\alpha^2-\beta^2)}} \approx e^{jkz} e^{-\frac{jkz}{2}(\alpha^2+\beta^2)}. \tag{3.36}$$

where k_x/k and k_y/k are the direction cosines in the transverse direction. Equation (3.36) is the transfer function for a plane wave in homogeneous, isotropic media. The term $\exp(jkz)$ is generally dropped although it may be important in phased array and coupled resonator studies.

We can associate the direction cosines with spatial frequency variables ξ and η,

$$\xi\lambda = \alpha \text{ and } \eta\lambda = \beta, \tag{3.37}$$

and the transfer function for a plane wave described in terms of spatial frequency variables is

$$e^{jk_z z} \approx e^{-j\pi\lambda z\rho^2}, \tag{3.38}$$

where $\rho^2 = \xi^2 + \eta^2$.

Any well behaved function may be written as a summation of spatial frequency components,

$$a(x, y; z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(\xi, \eta; 0) e^{j2\pi(x\xi + y\eta)} d\xi d\eta, \tag{3.39}$$

and from Eq. (3.38)

$$A(\xi, \eta; z) = A(\xi, \eta; 0) e^{-j\pi\lambda z\rho^2}. \tag{3.40}$$

Harvey has extended the scalar theory to work at large angle by taking good advantage of a direction cosine (rather than a small angle) development[5–6].

Using Eq. (3.40) propagation in homogeneous media can be written in the operator notation:

$$a(x, y; z) = \boldsymbol{FF}^{-1}[T(z)\boldsymbol{FF}[a(x, y; 0)]], \tag{3.41}$$

where,

$$T(z) = e^{-j\pi\lambda z\rho^2}. \tag{3.42}$$

is the transfer function of diffraction propagation. The forward and inverse Fourier transforms, $\boldsymbol{FF}$ and $\boldsymbol{FF}^{-1}$ are defined by,

$$\boldsymbol{FF}[\ ] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\ ] e^{-j2\pi(x\xi + y\eta)} dx dy, \tag{3.43}$$

$$\boldsymbol{FF}[\ ]^{-1} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\ ] e^{j2\pi(x\xi + y\eta)} d\xi d\eta. \tag{3.44}$$

Propagation may be written as a convolution by taking the Fourier transform of Eq. (3.40) [7].

$$a(x_2,y_2;z_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a(x_1,y_1;z_1)t(x_1-x_2,y_1-y_2,z_2-z_1)dx_1dy_1\,, \tag{3.45}$$

$$t(x,y;\Delta z) = \frac{1}{j\lambda z}e^{j\frac{kr^2}{2\Delta z}}. \tag{3.46}$$

The quantity $t(x,y;\Delta z)$ is the point spread function (PSF) or impulse response function. Phase factors which are constant over the field have been dropped. The quadratic phase factor of Eq. (3.35) can be factored to give Eq. (3.47).

$$a(x_2,y_2;z_2) = \frac{1}{j\lambda z}q(r_2,\Delta z)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a(x_1,y_1;z_1)q(r_1;\Delta z)e^{-j\frac{2\pi}{\lambda z}(x_1x_2+y_1y_2)}dx_1dy_1\,, \tag{3.47}$$

where $q(r;z) = \exp[jk(r^2/2z)]$ is a quadratic phase factor and simplifies many of the diffraction equations. In operator notation,

$$a(x_2,y_2;z_2) = \frac{1}{j\lambda\Delta z}q(r_2;\Delta z)\boldsymbol{FF}^s[a(x_1,y_1;z_1)q(r_1;\Delta z)]\,, \tag{3.48}$$

where $s = \Delta z/|\Delta z|$.

Equations (3.41) and (3.48) are the near field and far field propagation expressions. In the continuous mathematical formulation, there is no difference between the two expressions. In discrete formulation for numerical calculations, errors are reduced if the correct selection of a near or far field propagator is made. This arises from the quadratic phase factors that must be evaluated. In the near field, the phase factor is found from Eq. (3.42) and, in the far field the phase factor is found from Eq. (3.46):

$$T(\Delta z) = e^{-j\pi\lambda\Delta z\rho^2} \text{ and } t(x,y;\Delta z) = \frac{1}{j\lambda\Delta z}e^{j\frac{kr^2}{2\Delta z}}. \tag{3.49}$$

The phase factor, $T(\Delta z)$ for the near field phase factor varies rapidly as $\Delta z \to \infty$ but slowly as $\Delta z \to 0$. However, $t(\Delta z)$ varies slowly as $\Delta z \to \infty$ and rapidly as $\Delta z \to 0$. Rapidly varying phase factors create numerical errors called aliasing, which are described in more detail later in this manual. The near field propagator aliases at large propagation distances but is well behaved at short propagation distances.These relationships are summarized in Table 3.1.

By using the far field expression at long propagation distances and the near field expression at short distances, aliasing can be reduced to a tolerable level in most cases.

References

1. Sargent, Scully, and Lamb, *Laser Physics*, Addison Wesley, 1974.

Table. 3.1. Variation of quadratic phase factor for the two propagation algorithms for different distances.

	far-field	near-field
$\Delta z \to 0$	rapid	slow
$\Delta z \to \infty$	slow	rapid

2. N. Bloembergen, *Nonlinear Optics*, W. A. Benjamin, Inc., Reading, MA (1965).
3. R. H. Hardin and F. D. Tappert, "Applications of the split step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations," SIAM Rev. 15, 423 (1973).
4. J. W. Goodman, *Introduction to Fourier Optics*, McGraw Hill (1968).
5. J. Harvey, 'Perspectives on Diffraction," OE Magazine, 22–25 (2001).
6. J. Harvey and C. Vernold, et. al., Appl. Opt. **38**, 6469-6491, (Nov. 1, 1999).
7. H. Kraus, "Huygens Fresnel Kirchoff wave front diffraction formulation: spherical waves," J. Opt. Soc. Am. A, Vol. 6, No. 8, 1196 (1989).

4. Diffraction Phenomenology

It is possible to develop certain rules-of-thumb to characterize diffraction effects. One of the most useful concepts is the Fresnel number which may be used to characterize the diffraction of circular apertures. We shall also consider the concepts of near- and far-field corresponding to Fresnel and Fraunhofer regimes. Gaussian beams are an important special case because of their prevalence in laser systems. The Rayleigh range of gaussian beams serves as an alternate to the concepts of near- and far-field regions. Linear gratings have special significance for the angular spectrum theory and illustrate the concept of Talbot imaging and the characteristic diffraction length. The important relationship between the magnitude of phase aberrations and the presence of higher order harmonics in the far-field will be discussed. The concept of Huygen's wavelets proves to be very useful in developing a conceptual understanding of diffraction. We envision each point on the wavefront as being the source of spherical wavelets. The wavelets spread out at the speed of light in the medium. At a finite distance, the combination of all wavelets cancels except at the boundary which is locally parallel to the initial wavefront. A plane wavefront will retain its identity as a plane wavefront during propagation, as shown in Fig. 4.1. Similar results apply for a spherically diverging wavefront. The radius of curvature expands such that the spherical wavefront always has the same center-of-curvature. The effect at any distant point is the cumulative effect of the radiation from all the Huygen's wavelets.

Properly, the integration of these effects should be done along the spherical surface of the wavefront but more commonly the integration is taken over a plane, as was done in Section 3. Kraus has investigated some of the errors in the calculation of diffraction for spherical beams [7]. Consider a source point for a spherical wavefront and an observation point, as shown in Fig. 4.2. The optical path difference (OPD) is the optical path length (OPL) minus the axial distance,

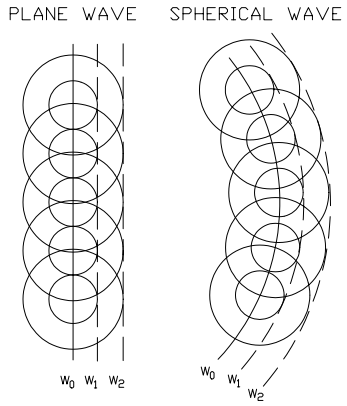


Fig. 4.1. Huygen's wavelets expanding from a plane wave (left) regenerate a plane wave. Huygen's wavelets from an expanding wave will regenerate an expanding wave.

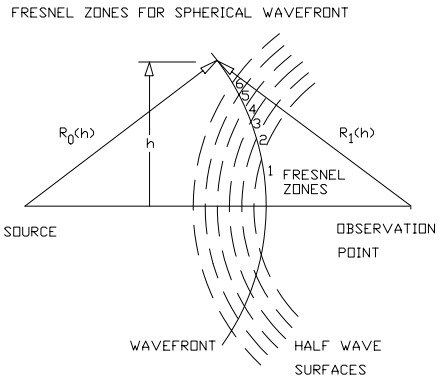


Fig. 4.2. Fresnel zones for a spherical wavefront.

$$OPL = \frac{2\pi}{\lambda}(R_0(h) - R_1(h)). \tag{4.1}$$

where h is the transverse radius. The optical path difference (OPD) is approximately,

$$OPD = \frac{\pi}{\lambda}h^2\left(\frac{1}{R_0(0)} - \frac{1}{R_1(0)}\right). \tag{4.2}$$

These regions are indicated by concentric circles centered at the observation point. The Fresnel zones are found by dividing the aperture up into regions whose boundaries are one-half period apart. The zones are numbered in Fig. 4.2. Odd numbered and even numbered zones add coherently separately. The variation of phase is, of course, continuous but it is easier to illustrate with discrete bands. The number of Fresnel zones is the Fresnel number. Fresnel numbers which are exactly even integers will result in zero intensity on axis. Odd integer values of the Fresnel number will result in a peak value of beam intensity.

It is more convenient to define the Fresnel number in terms of the radius of curvature of the wavefront. Let the radius of curvature of the wavefront be Z. Z is defined to be the distance from the center-of-curvature to the current position of the wavefront. We make the substitution,

$$R_o(0) = Z_o,\ R_1(0) = Z_0 - Z_1, \tag{4.3}$$

where the geometric radius of curvature increases linearly with the propagation of the wavefront in the positive z-direction, as shown in Fig. 4.3,

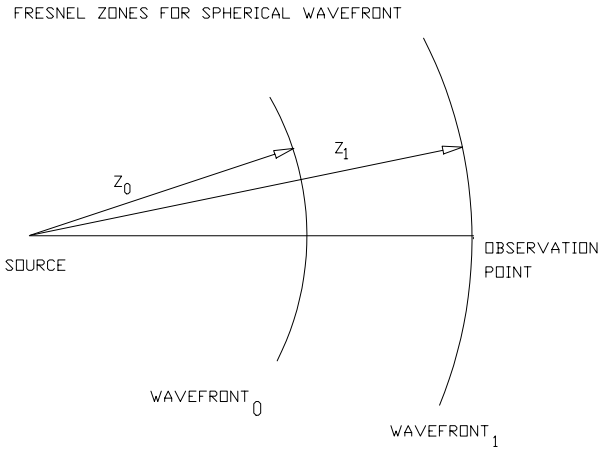


Fig. 4.3. Radii of expanding wavefront.

$$Z_1 = Z_0 + \Delta Z. \tag{4.4}$$

In terms of the radius of the phase curvature, the Fresnel number is

$$F_n = \frac{a^2}{\lambda Z_{eff}}, \tag{4.5}$$

where a is the beam radius

$$Z_{eff} = (Z_1 - Z_0)\frac{Z_0}{Z_1}, \tag{4.6}$$

where Z_{eff} is equivalent distance in collimated space. All beams having the same Fresnel number will have the same appearance apart from scale factors.

For an off-axis observation point, the half-wave phase surfaces are shifted with respect to the aperture. The Fresnel zones do not intercept the aperture in a symmetrical manner. The off-axis point will never see complete cancellation of the light for an even Fresnel number. The peak value for an odd Fresnel number will be lower than for an on-axis point.

Where clear apertures are well-defined and the beam is of nearly uniform irradiance, the Fresnel number may be used to specify the condition of the beam with respect to near- or far-field status. A useful rule-of-thumb for whether the function is in the near- or far-field is

- near-field, if $\infty \geq F_n \geq 1$
- far-field, if $1 \geq F_n \geq 0$

These expressions may be applied separately to the aperture sizes and degree of collimation for the x- and y-directions. The Fresnel number is useful for simple systems with a single well-defined aperture. This is always the case in geometrical optics, but in physical optics analysis there may be no apertures of significance or several, as shown in Fig. 4.4.

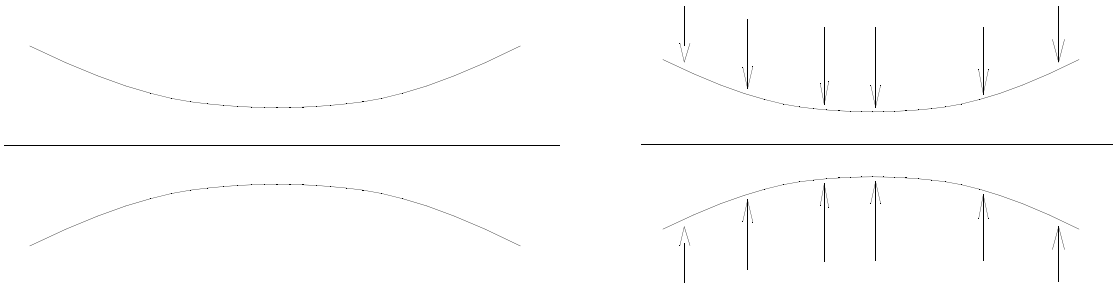


Fig. 4.4. Some systems have no apertures (left). Others have many (right). Fresnel number concepts deal poorly with either type of system.

For a circular aperture the diffraction point spread function of Eq. (3.46) can be expressed in terms of the Fresnel number,

$$t(x,y) = \frac{1}{j\lambda z} e^{\frac{j\pi F_n r^2}{a^2}}. \tag{4.7}$$

We want to know the spatial frequency exhibited at the point x. Taking the derivative of the quadratic phase factor, we have

$$\frac{d\theta}{dr} = 2\pi F_n \frac{r}{a^2}, \tag{4.8}$$

where θ is the phase of the quadratic phase factor. This corresponds to a spatial frequency of

$$f = F_n \frac{r}{a^2}. \tag{4.9}$$

The spatial frequency in the point spread function (PSF) at the edges of the aperture determine the spatial frequencies present in the diffraction pattern. If r_1 and r_2 are the distances to the edges from the point of observation, then the frequencies are

$$f_1 = F_n \frac{r_1}{a^2} \text{ and } f_2 = F_n \frac{r_2}{a^2}. \tag{4.10}$$

Figure 4.5 shows schematically the radii from the observation point to the edges of the aperture.

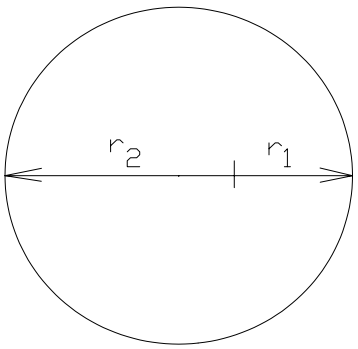


Fig. 4.5. Schematic of the distances from the observation point to the edges of the aperture. The frequencies present in diffraction pattern depend linearly on the distances r_1 and r_2.

Figure 4.6 shows how the quadratic phase factor appears in the aperture with the distribution centered, slightly displaced (P_1), more displaced (P_2), and with the distribution displaced by a from the center of the aperture. The displacement of the quadratic phase factor gives the point on the near-field, free-space diffraction function.

Near the center the spatial frequency of the intensity modulation is approximately F_n/a. The degree of modulation is strong because all parts of the aperture interact with the same spatial frequency in the PSF. We can see that going from the centered condition to Point P_1 results in a 1/2 cycle oscillation. The outer dark band is partially blocked on the right and a white band is partially illuminated. At the center, both edges give essentially the same frequencies. At the geometric shadow, the slight ripple in the diffraction pattern

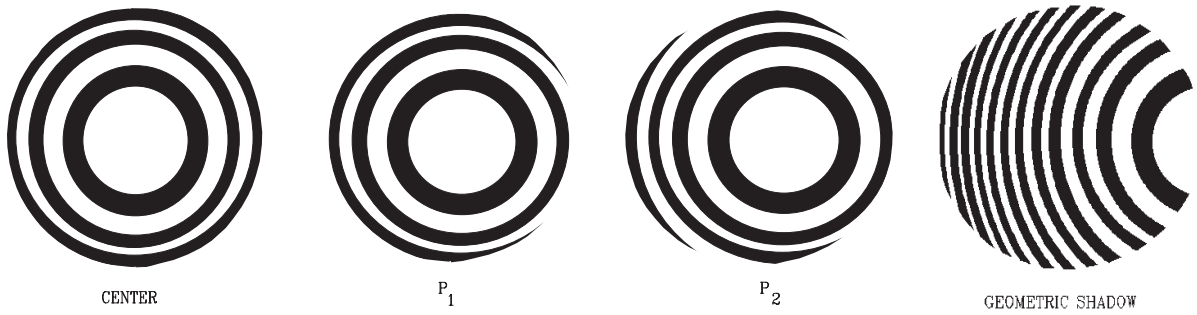


Fig. 4.6. Appearance of the quadratic phase factor with the observation point at the center, slightly shifted, and shifted o correspond to the geometric edge of the aperture. The principle source of variations in intensity is the blocking or exposure of regions of constructively or destructive interference.

comes from the opposite edge of the aperture at twice the frequency in the center. The typical frequencies at the center and the edge are

$$f_{center} = \frac{F_n}{a} \text{ and } f_{edge} = \frac{2F_n}{a}. \tag{4.11}$$

As shown in Fig. 4.7, at the center of the diffraction pattern, both radii are the same $r_1 = r_2 = a$. At intermediate points, two frequencies are present and the diffraction pattern shows the effects of these two frequencies beating together. At the edge, one frequency goes to zero and the other goes to $r = 2a$. At the edge, the distribution falls of sharply when one frequency goes to zero, but also shows a small amount of oscillation in the shadow region.

Figure 4.8 show profiles of the intensity in a diffraction pattern of a circular aperture for Fresnel numbers 5 and 6. Fresnel number 5 shows the peak intensity at 4 which is expected from a uniform intensity of 1 across the aperture. Note that there are 5 principle bumps across the aperture. The peak intensity of $4I_0$, where I_0 is the initial intensity, is characteristic of odd Fresnel numbers.

Fresnel number 6 shows a center intensity of zero typical of even Fresnel numbers. There are 6 major bumps across the aperture the same as the Fresnel number. Both patterns show 25 percent intensity at the geometric shadow. The extra bumps are due to beating effects between the frequencies generated from diffraction from the two edges. The very fine bumps outside the geometric shadow are due to diffraction from the far edge of the diffracting aperture.

4.1 Single Plane Wave Component

Diffraction alters the complex amplitude distribution with the higher frequency components being most affected. Plane waves, $e^{j\mathbf{k}\cdot\mathbf{r}}$, are eigenfunctions of free-space propagation. For small angles, plane waves propagate as described in Eq. (3.43),

$$\varphi(\Delta z) = \varphi(0)e^{-j\pi\lambda\Delta z(f_x^2+f_y^2)}. \tag{4.12}$$

where the leading phase factor $e^{jk\Delta z}$ has been dropped. The phase of the phasor is

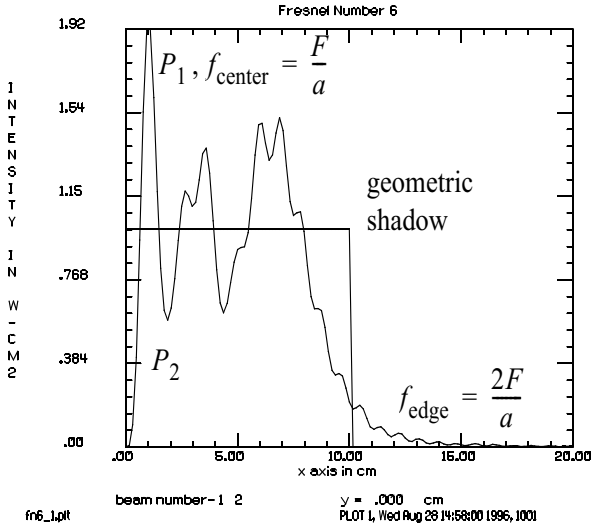


Fig. 4.7. The diffraction pattern of a flat top function exhibits characteristic frequencies. Near the center $f \approx F_n / A = a$. In the region of the shadow, $f \approx 2F_n / a$. The center, P_1, P_2, and the edge relate roughly to the Fresnel zones illustrated in Figs. 4.8.

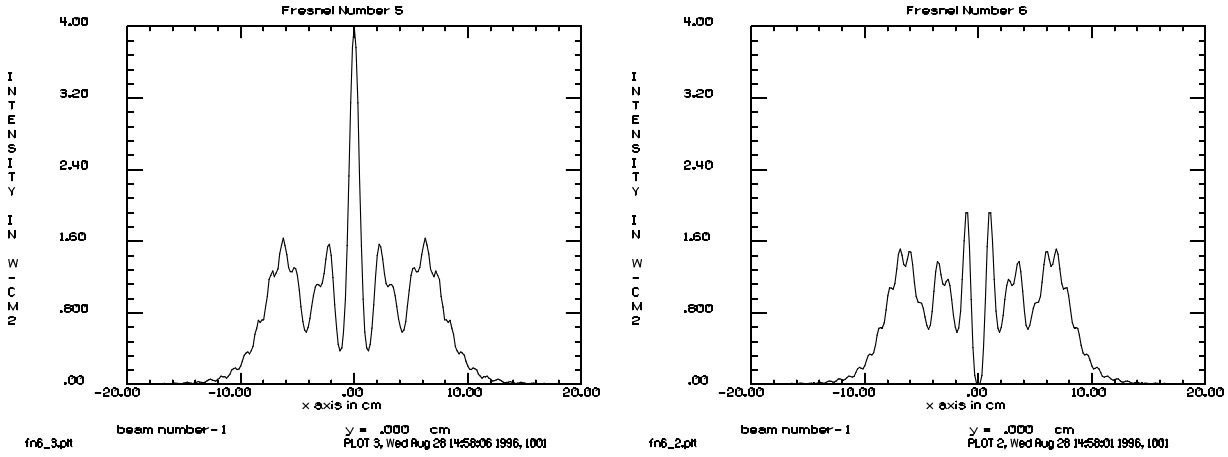


Fig. 4.8. Fresnel numbers 5 (left) and 6 (right). The extra bumps are due to beating effects between the frequencies generated from diffraction from the two edges. The very fine bumps outside the geometric shadow are due to diffraction from the far edge of the diffracting aperture.

$$\theta = -\pi\lambda\Delta z(f_x^2 + f_y^2). \tag{4.13}$$

The frequency phasor will move cyclically from

1) pure real, $\theta = 0$

2) pure imaginary, $\theta = \frac{\pi}{2}$

3) pure real, reversed phase, $\theta = \pi$

4) pure imaginary, reversed phase, $3\frac{\pi}{2}$

5) pure real, $\theta = 2\pi$

The cyclical effect has an axial spatial period of $\Delta z = 2/\lambda f^2$. The frequency component is exactly reproduced every full period. This effect was first reported by Talbot in 1836 and explained mathematically in 1881 by Lord Rayleigh [4.1,4.2]. It is illustrated schematically in Fig. 4.9. The characteristic diffraction length can be defined to be a quarter cycle of the Talbot imaging period.

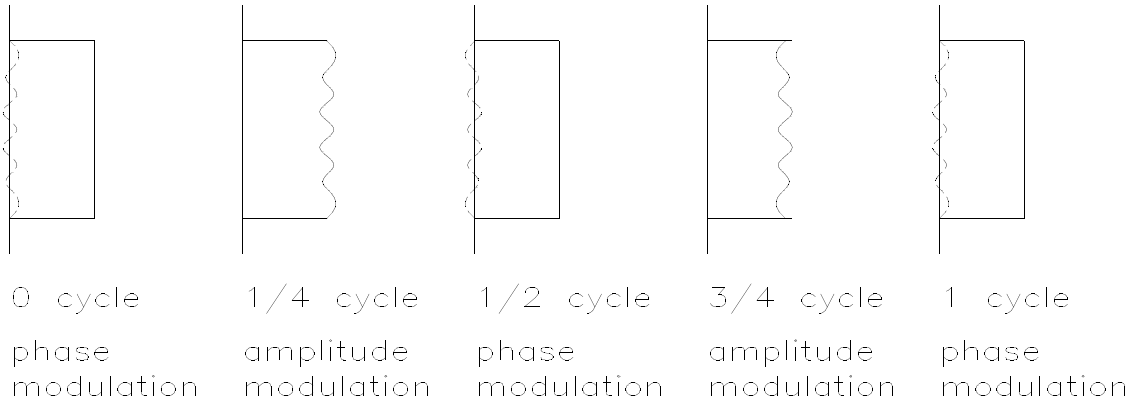


Fig. 4.9. at small periodic modulation in amplitude or phase will behave cyclically.

$$\Delta z_{char} = \frac{1}{2\lambda f^2}. \tag{4.14}$$

Propagation of one characteristic length will result in conversion of amplitude modulation to phase modulation and vice versa. Calculations of diffraction and nonlinear gain require that the axis be sampled at least twice per characteristic length to get the correct amplification of the high frequency amplitude components which are appearing and disappearing during propagation. The amplitude and phase of a phase grating at the initial condition and after one-quarter Talbot cycles are shown in Fig. 4.10.a and 4.10b.

If the function $\psi(x,y)$ has phase aberrations, we can analyze their effects by considering each frequency component of the aberration. Let $\psi(x,y)$ be

$$\psi(x,y) = A(x,y)e^{jkW(x,y)}. \tag{4.15}$$

Consider a complex amplitude distribution with a single spatial frequency in the wavefront

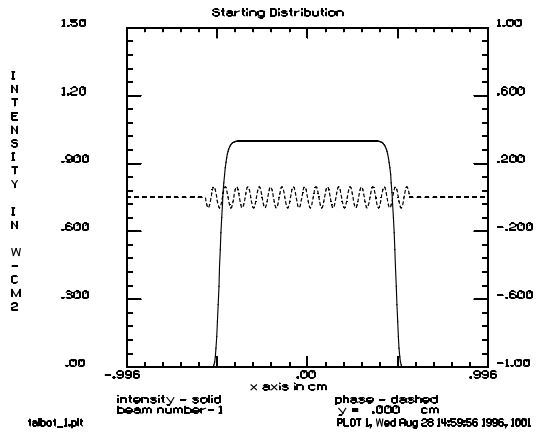


Fig. 4.10.a. Initially, the amplitude is selected to be a top hat function and sinusoidal phase modulation is imposed. The effect may be observed for small phase modulation if the spatial frequency is sufficiently high.

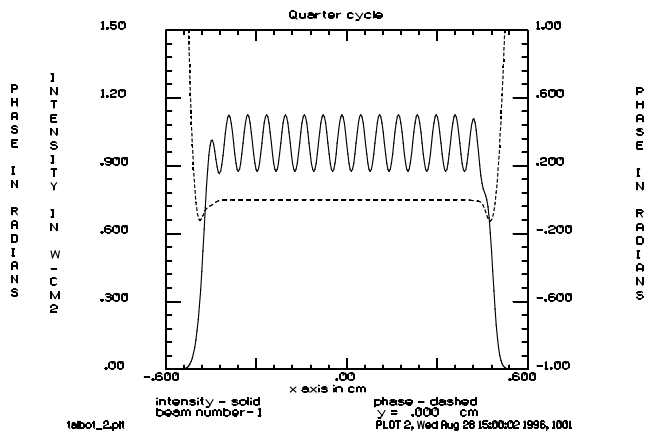


Fig. 4.10b. After diffraction of one-quarter Talbot cycle, the phase modulation has turned to amplitude modulation. Note, that the phase modulation is gone except in the regions near the edge of the aperture.

$$\psi(x) = e^{jC\sin(2\pi fx)} \tag{4.16}$$

where C is the amplitude of the wavefront modulation. We can identify the frequency components, from a Bessel function relationship

$$a(x) = \sum_{n=-\infty}^{\infty} J_n(C)e^{j2\pi nfx}. \tag{4.17}$$

The spatial frequencies are

$$A(\xi) = \sum_{n=-\infty}^{\infty} J_n(C)\delta(\xi - nf). \tag{4.18}$$

As can be seen from Eq. (4.18), a single spatial frequency in phase has a range of frequency harmonics in complex amplitude. When the phase modulation is small, we may consider only the lowest harmonic frequency. For larger aberration coefficients, higher order harmonics must be considered. This may significantly increase the sampling requirements for proper representation because we must represent the complex amplitude in the array not the amplitude and wavefront.

4.2 Far-Field Diffraction Effects

The far-field diffraction pattern is formed either by propagating a great distance or by propagating to the focal point of a lens. Far-field diffraction is described by Fraunhofer diffraction. The complex amplitude in the image plane is

$$a(x',y') = \frac{1}{j\lambda f} e^{\frac{jkr'^2}{2f}} \boldsymbol{FF}[a(x,y)], \tag{4.19}$$

$$a(x',y') = \frac{1}{j\lambda f} e^{\frac{jkr^2}{2f}} \iint a(x,y) e^{-j\frac{2\pi}{\lambda f}(xx'+yy')} dxdy, \tag{4.20}$$

where

- $a(x,y)$ is the amplitude distribution in the pupil
- $a(x',y')$ is the amplitude distribution in the final plane of observation
- f is the focal length of the lens
- $\boldsymbol{FF}$ is a two-dimensional Fourier transform from Eq. (3.43) with $\xi = \frac{x'}{\lambda f}$, $\eta = \frac{y'}{\lambda f}$.

The quadratic phase factor makes the complex image distribution not exactly the Fourier transform of the pupil distribution. If we first propagate a distance f prior to the lens, the quadratic phase factor will be deleted. Propagation a distance f in the near-field imposes a quadratic phase factor in frequency space of

$$H(f) = e^{-j\pi\lambda f\rho^2} = e^{\frac{-jkr^2}{2f}}, \tag{4.21}$$

exactly the correct amount to cancel the quadratic phase factor of Eq. (4.19). We can see that a simple lens forms the exact Fourier transform of whatever is in the front focal plane of the lens. If we are only concerned about the intensity in the far-field, we can disregard the quadratic phase factor and claim that the lens forms a Fourier transform of the pupil.

For a uniformly filled pupil, the far-field pattern is

$$a(x',y') = \frac{a^2}{jkz} e^{jk\frac{r^2}{2f}} \frac{J_1\left(\frac{2\pi ar'}{\lambda f}\right)}{\frac{ar'}{\lambda f}}, \tag{4.22}$$

where J_1 is the first order Bessel function of the first kind and a is the aperture radius. We have dropped a factor of e^{jkz}, a constant phase term.

The irradiance distribution is

$$i(x',y') = |a(x',y')|^2 = \frac{a^4}{\lambda^2 f^2}\left(\frac{J_1\left(\frac{2\pi ar'}{\lambda f}\right)}{\frac{ar'}{\lambda f}}\right)^2. \tag{4.23}$$

This is the well-known Airy pattern.

4.3 Gaussian Beams

With the advent of the laser, great interest has been generated in gaussian beams. Well-behaved stable laser systems generate beams which are very nearly gaussian beams[3]. More recently Seward has argued that the equations may be written more simply in terms of diameters and full angles[4], but we will use the conventional treatment in terms of radii and half angles. The gaussian beam is the lowest mode of a set of Hermite gaussian beams. The Hermite gaussians are orthogonal and complete so they are convenient for describing arbitrary functions. All of the Hermite-gaussian functions can be propagated analytically, so they provide an alternate method to the Fourier transform propagation techniques.

The gaussian beam has several interesting properties

- analytically propagatable
- its own Fourier transform
- ideally smooth
- tightest focus position is not at the paraxial focus

The gaussian beam is characterized by a well-defined minimum beam size called the waist. The complex amplitude at the waist is of the form[3]

$$a(r,0) = e^{-\frac{r^2}{\omega_o^2}}, \tag{4.24}$$

where r is the transverse coordinate and ω_o is the gaussian transverse radius at the waist.

The definition of the gaussian spherical wave adds a quadratic phase factor:

$$a(r,0) = e^{-\frac{r^2}{\omega_o^2}} e^{-\frac{jkr^2}{R}}, \tag{4.25}$$

where k is the wave number and R is the radius of the phasefront. At the waist $R = \infty$, and the phase factor disappears.

To the accuracy of scalar Fresnel diffraction, the gaussian beam propagates according to the equation.

$$a(r,0) = e^{j(kz-\theta(z))} e^{-\frac{r^2}{\omega_o^2}} e^{-\frac{jkr^2}{R}}, \tag{4.26}$$

where we have used the normalized form. The amplitude drops proportionally to the increase in beam radius. There is a piston term which contains a phase term specific to gaussian beams. The remainder of the function is a spherical gaussian wave. For a traveling wave coordinate system we drop the $\exp(jkz)$ term because we assume a traveling coordinate system moving at the speed of light:

$$\omega(z) = \omega_o \sqrt{1 + \left(\frac{z}{z_R}\right)^2} \text{ radius of beam,} \tag{4.27}$$

$$R(z) = z + \frac{z_R^2}{z} \text{ phase radius,} \tag{4.28}$$

$$\theta(z) = \tan^{-1}\left(\frac{z}{z_R}\right) \text{ phase factor, Gouy shift,} \tag{4.29}$$

$$z_R = \frac{\pi\omega_o^2}{\lambda} \text{ Rayleigh distance.} \tag{4.30}$$

4.4 Significance of Rayleigh distance

The transverse waist radius increases away from the waist to form a hyperbolic surface of revolution. The phase term $\theta(z)$ is a piston term varying from $\pi/2$ to $-\pi/2$ when propagating from $z = \infty$ to $z = -\infty$. The local transverse distribution is scaled according to the local beam radius, $\omega(z)$.The radius of the quadratic phase factor, $R(z)$, varies from ∞ at the waist to a minimum value of $2z_R$ at the Rayleigh distance, and then back to ∞ when $z = \infty$. The maximum curvature of the wavefront occurs at the Rayleigh distance. The beam size has increased by $\sqrt{2}$ at the Rayleigh distance. The spherical harmonic wave transforms itself into another spherical harmonic wave upon propagation. This is not strictly an eigenfunction relationship. The eigenfunction of free-space is the plane wave. However, the gaussian spherical harmonic wave is almost an eigenfunction.

4.4.1 Finding the Waist

Given the spherical gaussian wave properties at some arbitrary point, we can calculate the location and size of the waist. Given the beam waist size, $\omega(z)$, and the phase radius, $R(z)$, we seek a value of z and z_R which will satisfy the following equations:

$$R(z) = z + \frac{z_R^2}{z},\ \omega(z) = \omega_o\sqrt{1 + \left(\frac{z}{z_R}\right)^2}. \tag{4.31}$$

The solution to these equations is

$$z = -\frac{R(z)}{1 + \left(\frac{\lambda R}{\pi\omega(z)^2}\right)^2},\ \omega_o = \frac{\omega(z)}{\sqrt{1 + \left(\frac{\pi\omega(z)}{\lambda R}\right)^2}}. \tag{4.32}$$

4.4.2 Lens Law for Gaussian Beams

To determine the properties of a thin lens acting on a gaussian beam, we simply consider the local phase radius of curvature to be transformed according to,

$$\frac{1}{R_1} - \frac{1}{R_2} = \frac{1}{f} \tag{4.33}$$

Note that we have adopted a sign convention for gaussian beams for the radius to be positive if the center-of-curvature is to the left, opposite to the usual convention for geometrical optics but consistent with Siegman's treatment [3]. The lens simply modifies the phasor radius of curvature. The beam size is unmodified. If we need to calculate aberrations of the lens, we should use the local radii of curvature to define the object and image conjugates.

The Rayleigh distance may be used to develop a concept of near- and far-field which is a generally more useful definition. For any beam we can find a gaussian beam which best fits the distribution. We can then consider whether the (best fit gaussian beam is inside or outside the Rayleigh range.

4.5 Hermite Gaussian Beams

The transverse modes of ideal stable resonators take the form of Hermite gaussian polynomials or Laguerre gaussian. The general polynomial form of the Hermite gaussian functions is

$$u_n(x) = \left(\frac{2}{\pi}\right)^{1/4}\left(\frac{1}{2^n n!\omega_o}\right)^{1/2} H_n\left(\frac{\sqrt{2}x}{\omega(z)}\right) e^{-\frac{x^2}{\omega^2}} \tag{4.34}$$

where n is the order of the polynomial, ω_o is a waist radius parameter similar to the gaussian beam and $H_n(x)$ are the Hermite functions. The two-dimensional functions may be described by multiplying two one-dimensional functions. The order and waist parameters may be different for the two directions.

The Hermite gaussian functions may be propagated analytically. Let q_1 and z_1 be the initial q-factor and axial position, where

$$q_1 = \left(\frac{1}{R_1} - \frac{j\lambda}{\pi\omega_1}\right)^{-1}. \tag{4.35}$$

For a propagation to the new plane z, the new q-parameter is

$$q_1 = \left(\frac{1}{R_1} - \frac{j\lambda}{\pi\omega_1^2}\right)^{-1} = q_1 + z - z_1. \tag{4.36}$$

We further define the parameter $Q(z)$, such that

$$Q(z) = \frac{q(z)}{q_1}. \tag{4.37}$$

The Hermite gaussian function at plane z is

$$f_n(x,y) = \left(\frac{2}{\pi}\right)^{1/4}\left(\frac{1}{2^n n!\omega_1 Q(z)}\right)^{1/2}\left(\frac{Q^*(z)}{Q(z)}\right)^{n/2} H_n\left(2^{1/2}\frac{x}{\omega}\right)\exp\left[-j\left(\frac{kx^2}{2q(z)}\right)\right]. \tag{4.38}$$

Both the Hermite polynomial H_n in the numerator and the term $2^n n!$ in the denominator become very large, even though their ratio is well behaved. The Hermite gaussian may be directly calculated recursively to achieve better numerical accuracy and less risk of having intermediate values exceed the word length of the computer (See Ref. 5),

$$f_o(x,z) = \left(\frac{2}{\pi}\right)^{1/4}\left(\frac{1}{\omega_1 Q(z)}\right)^{1/2}\exp\left[-j\left(\frac{kx^2}{2q(z)}\right)\right], \tag{4.39}$$

$$f_1(x) = \frac{2x}{\omega} f_o(x). \tag{4.40}$$

The higher order terms may be generated recursively by,

$$f_n(x,z) = \left(\frac{Q^*(z)}{nQ(z)}\right)^{1/2}\frac{2x}{\omega(z)} f_{n-1}(x,z) - \left(\frac{n-1}{n}\right)^{1/2}\frac{Q^*(z)}{Q(z)} f_{n-2}(x,y). \tag{4.41}$$

In the special case where $z = z_1$, i.e., no propagation, and evaluation of the Hermite gaussian at the waist; we have $R(z) = \infty$ and $Q = Q^* = 1$:

$$f_o(x) = \left(\frac{2}{\pi}\right)^{1/4}\frac{1}{\omega^{1/2}}\exp\left(-\frac{x^2}{\omega^2}\right), \tag{4.42}$$

$$f_1(x) = \frac{2x}{\omega} f_o(x). \tag{4.43}$$

The higher order terms may be generated recursively by,

$$f_n(x) = \frac{1}{n^{1/2}}\frac{2x}{\omega} f_{n-1}(x) - \left(\frac{n-1}{n}\right)^{1/2} f_{n-2}(x). \tag{4.44}$$

Calculation of the Hermite gaussian functions using Eqs. 4.42, 4.43, and 4.44, appears to avoid the problem of large numbers when evaluating Eq. 4.34 directly. The speed of calculation should be essentially the same as for H_n, provided a table of $n^{1/2}$ is precalculated.

References

1. H. F. Talbot, “Facts relating to Optical Science. No. IV.,” Phil. Mag. (Third Series), 9, p401 (1836).

2. Lord Rayleigh, “On copying Diffraction-gratings, and on some Phenomena connected therewith,” Phil. Mag. (Fifth Series), 11, p196 (1881).
3. A. E. Siegman, *Lasers*, University science Books, Mill Valley,CA(1986).
4. George H. Steward, “Two-dimensional space-angle product of a Gaussian beam,” Opt. Eng. **40**(9) 1959–1962 (2001).
5. A. E. Siegman and Edward A. Sziklas, “Mode Calculation in Unstable Resonators with Flowing Saturable Gain. 1:Hermite-Gaussian Expansion,” Appl. Opt., Vol. 13, p2775-2791 (1974).

5. Propagation in Homogeneous Media

The propagation through any well-behaved system can be separated into geometrical aberration calculations and propagation in homogeneous media. This concept will be more thoroughly developed in the next section on optical systems. In this section the methods for propagation — the hardest part of the modeling problem — will be described. The mathematically equivalent expressions of Eqs. (4.18) and (4.19) provide a complete description of diffraction propagation in homogeneous media in the Fresnel approximation. These equations form the basis of a set of numerical procedures which can treat all possible problems of propagation in homogeneous media. There are three major issues to consider:

- sampling interval,
- relative filling of the computer array,
- use of reference surfaces.

The sampling interval will determine the highest spatial frequency which can be represented. The relative size of the distribution in the computer array determines the amount of amplitude which falls outside the array boundaries and which folds back into the array to cause aliasing. When dealing with noncollimated beams, it is often advantageous to use curved reference surfaces to minimize the magnitude of phase variations that must be included in the complex amplitude distribution. By removing all or most of the phase associated with divergence or convergence, the array need only contain the phase which represents aberration of the beam. The general principles of sampled representation and discrete Fourier transforms will be developed. Sampling guidelines based on knowledge of the characteristic spatial frequencies in the diffraction pattern will be developed. The nature of aliasing and guidelines for the maximum allowable filling of the computer array will also be explored. Last, a consistent set of algorithms will be presented to treat all cases of diffraction propagation.The development will be done in separable form.

In numerical calculations, only discrete points may be represented. Also, only a limited region of space may be considered because of computer memory limitations. Consider a two-dimensional function represented in a rectangular computer array of $M \times N$ points as shown in Fig. 5.1. The sampling intervals for the x- and y-directions are Δx and Δy. In the general case, $M \neq N$ and $x \neq y$. The width of the computer array representation is $M\Delta x$ by $N\Delta y$. Information exists in the computer only at the discrete points defined by the rectangular grid. Any functions to be represented must be truncated by the finite width of the computer array.

The computer points in the spatial domain will be counted with the indices k and l. The indices have the ranges

$$\frac{-M}{2} \leq k \leq \frac{M}{2} - 1 \,, \quad \frac{-N}{2} \leq l \leq \frac{N}{2} - 1 \,. \tag{5.1}$$

Note that the center of the distribution has been chosen to be at $(M/2 + 1, N/2 + 1)$. Many Fast Fourier Transform (FFT) routines based on arrays dimensions which are powers of 2 are implemented with natural centers either at $(1,1)$ or $(M/2 + 1, N/2 + 1)$ by shifting the array one-half cycle in each direction. The

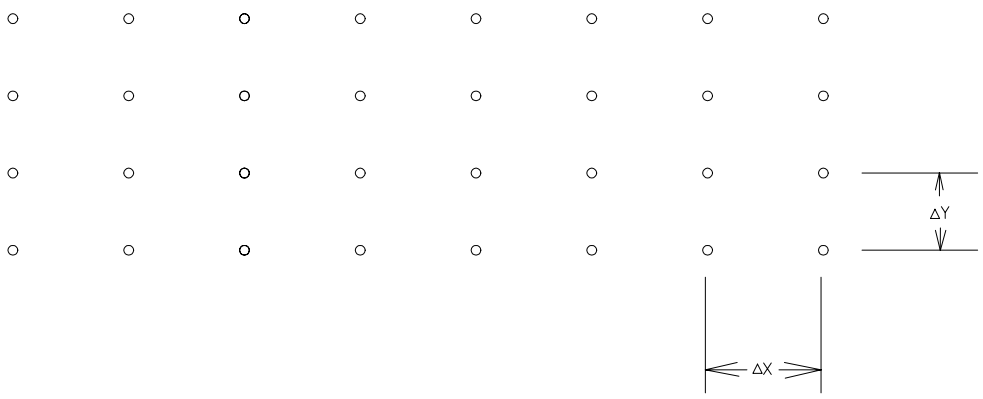


Fig. 5.1. Rectangular computer array representation of the function. In this case $M = 8$, $N = 4$, and Δx is greater than Δy.

natural center of the array is defined to be the point at which a delta function will give a perfectly constant real Fourier transform. An article by John Hayes in Applied Optics and Optical Engineering discusses the properties of FFT's in detail [1]. The physical limits are obtained by multiplying Eq. (5.1) by Δx and Δy,

$$-\frac{M\Delta x}{2} \le k\Delta x \le \frac{M\Delta x}{2} - \Delta x\,, \quad -\frac{N\Delta y}{2} \le l\Delta y \le \frac{N\Delta y}{2} - \Delta y\,. \tag{5.2}$$

Sampling can be represented as multiplication by a special function called the comb function. The comb function is an infinite array of delta functions spaced apart by Δx and Δy,

$$\mathrm{comb}\left(\frac{x}{\Delta x}, \frac{y}{\Delta y}\right) = |\Delta x||\Delta y| \sum_k \sum_l \delta(x - k\Delta x, y - l\Delta y)\,. \tag{5.3}$$

The comb function is useful in transforming a continuous function into a discrete representation,

$$a(x, y) \to a(x, y)\mathrm{comb}\left(\frac{x}{\Delta x}, \frac{y}{\Delta y}\right). \tag{5.4}$$

where $a(x, y)$ is the continuous function to be sampled. The arrow indicates transformation from continuous to discrete form.

The discrete nature of the spatial domain causes the frequency domain to be periodic (and necessarily of infinite extent). The continuous function $A(\xi, \eta)$, the Fourier transform of $a(x, y)$ is replicated with a period of $(1/\Delta x, 1/\Delta y)$. This is illustrated in Fig. 5.2.

The Fourier transform domain functions must also be discrete. The most common (and most efficient) form of the FFT has the same dimensions for the spatial and frequency domains. The frequency domain indices m and n have the ranges

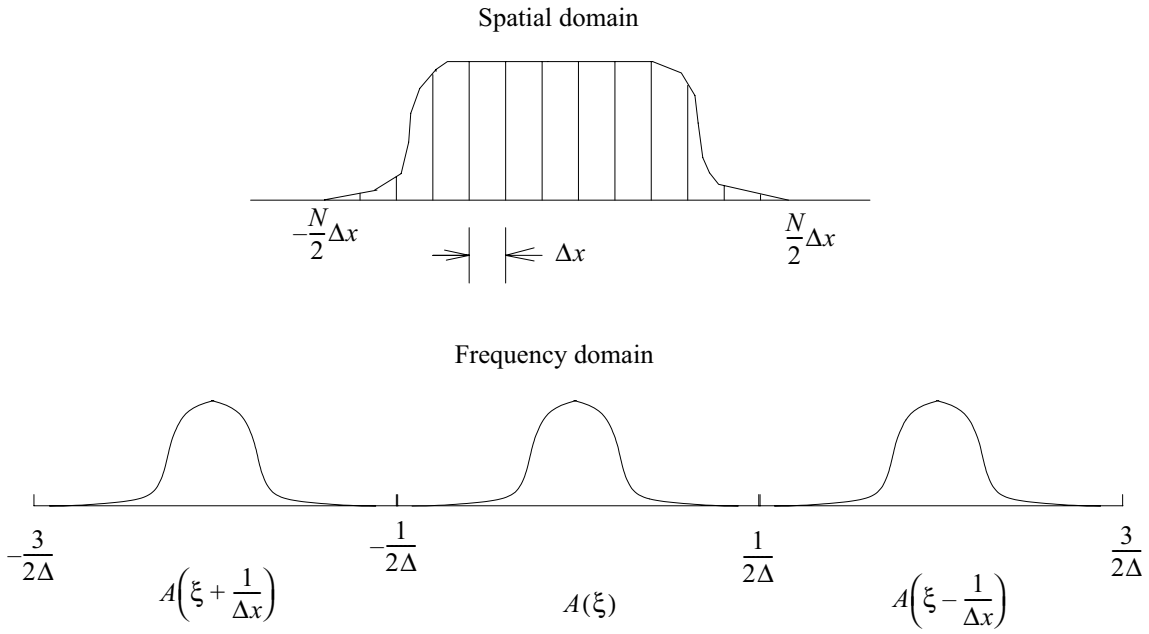


Fig. 5.2. Sampling of the spatial domain causes the frequency domain to be periodic with period $(1/\Delta x, 1/\Delta y)$.

$$-\frac{M}{2} \le m \le \frac{M}{2} - 1\,, \quad -\frac{N}{2} \le n \le \frac{N}{2} - 1\,. \tag{5.5}$$

Multiplication of Eq. (5.5) by $\Delta\xi = 1/(M\Delta x)$ and $\Delta\eta = 1/(N\Delta y)$ gives the frequency range,

$$-\frac{1}{2\Delta x} \le \xi \le \frac{1}{2\Delta x}\left(1 - \frac{2}{M}\right), \quad -\frac{1}{2\Delta y} \le \eta \le \frac{1}{2\Delta y}\left(1 - \frac{2}{N}\right). \tag{5.6}$$

The frequency domain bounds are the Nyquist sampling frequencies. The FFT algorithm is occasionally blamed for this restriction, but it is more accurately attributed to the discrete sampling process and will exist for any form of propagation of sampled data.

The continuous frequency domain function is also transformed to discrete representation by means of the comb function,

$$A(\xi, \eta) \to A(\xi, \eta)\mathrm{comb}\left(\frac{\xi}{\Delta\xi}, \frac{\eta}{\Delta\eta}\right) = A(m\Delta\xi, n\Delta\eta)\,. \tag{5.7}$$

The discrete nature of the frequency domain forces the spatial domain to also be periodic with period $(1/(\Delta\xi), 1/(\Delta\eta))$. These relationships are shown schematically in Fig. 5.3.

A Fourier transform pair can be defined for modified spatial and frequency functions $a(k\Delta x, l\Delta y)$ and $A(m\Delta\xi, n\Delta\eta)$, such that we have an exact Fourier relationship,

$$a(k\Delta x, l\Delta y) \Leftrightarrow A(m\Delta\xi, n\Delta\eta)\,, \tag{5.8a}$$

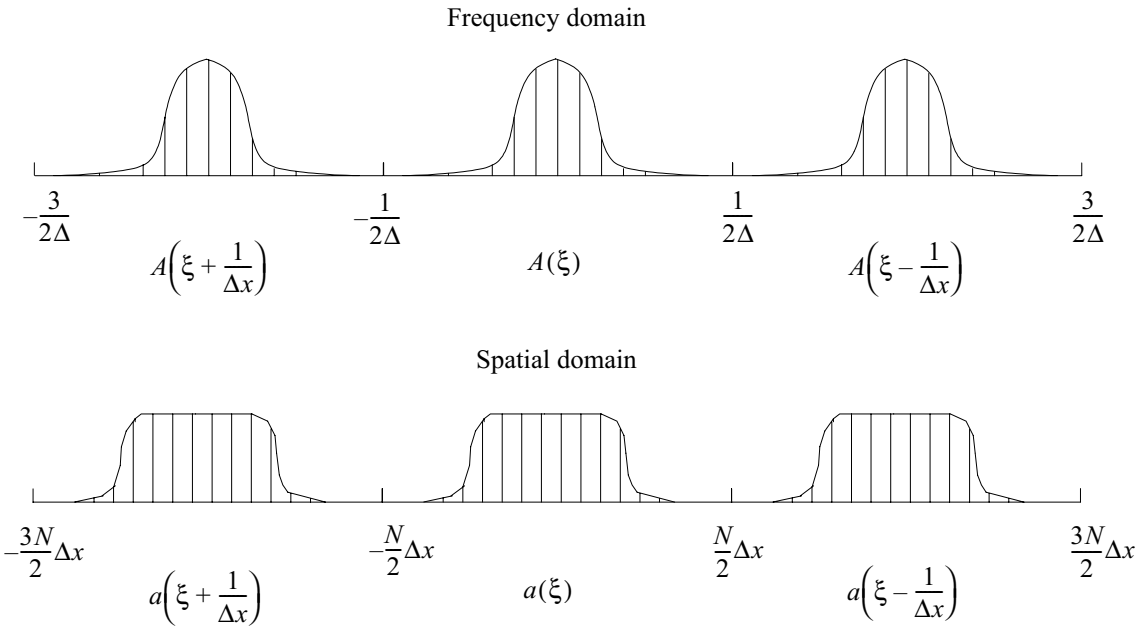


Fig. 5.3. Sampling periods in the spatial and frequency domains.

$$a(k\Delta x, l\Delta y) \Leftrightarrow a(x, y)\mathrm{comb}\left(\frac{x}{\Delta x}, \frac{y}{\Delta y}\right) ** |\Delta\xi\Delta\eta|\mathrm{comb}(x\Delta\xi, y\Delta\eta)\,, \tag{5.8b}$$

$$A(m\Delta\xi, n\Delta\eta) \Leftrightarrow A\mathrm{comb}\left(\frac{\xi}{\Delta\xi}, \frac{\eta}{\Delta\eta}\right) ** |\Delta x\Delta y|\mathrm{comb}(\xi\Delta x, \eta\Delta y)\,, \tag{5.8c}$$

where $**$ indicates two-dimensional convolution and $\Leftrightarrow$ indicates two-dimensional Fourier transformation pairs.

The function, $\mathrm{comb}(x\Delta\xi, y\Delta\eta)$, causes the spatial domain to be periodic with minimum periods of $M\Delta x$ and $N\Delta y$ in the x- and y-directions. There is, in effect, an infinite rectangular array of functions separated by $M\Delta x$ and $N\Delta y$. Therefore, the frequency domain sampling periods are

$$\Delta\xi = \frac{1}{M\Delta x}, \quad \Delta\eta = \frac{1}{N\Delta y}. \tag{5.9}$$

The Fourier transform operator can be written in discrete form,

$$\boldsymbol{FF}[\ ] = \sum_k \sum_l [\ ] \exp\left(-sj2\pi\left(\frac{km}{M} + \frac{ln}{N}\right)\right), \tag{5.10}$$

where s is +1 for forward transformation and -1 for inverse. Various forms of algorithms are used in FFT's and some have a normalization step for the forward or inverse transformation.

Evaluation of the far-field expression, Eq. (3.48), in discrete terms causes a redefinition of sampling period,

$$A(m\Delta\xi, n\Delta\eta) = \boldsymbol{FF}[a(k\Delta x, l\Delta y)], \tag{5.11a}$$

$$\Delta\xi = \frac{1}{M\Delta x_1}, \quad \Delta\eta = \frac{1}{N\Delta y_1}. \tag{5.11b}$$

The coordinates x_2, y_2 are related to $\Delta\xi, \Delta\eta$ by,

$$\Delta\xi = \frac{\Delta x_2}{\lambda|\Delta z|}, \quad \Delta\eta = \frac{\Delta y_2}{\lambda|\Delta z|}, \tag{5.12}$$

based on Eq. (3.48). The discrete far-field calculation is therefore

$$a(k\Delta z_2, l\Delta z_2) = \frac{1}{j\lambda\Delta z} q(r_2, \Delta z)\boldsymbol{FF}^s[a(k\Delta x_1, l\Delta y_1)q(r_1, \Delta z)], \tag{5.13a}$$

$$\Delta x_2 = \frac{\lambda|\Delta z|}{M\Delta x_1}, \quad \Delta y_2 = \frac{\lambda|\Delta z|}{N\Delta y_1}, \tag{5.13b}$$

$$r^2 = (k\Delta x)^2 + (l\Delta y)^2, \tag{5.13c}$$

$$s = \frac{z}{|\Delta z|}. \tag{5.13d}$$

Note, the scale change of the new sampling periods, Δx_2 and Δy_2. The discrete near-field propagation equation is

$$a(k\Delta x, l\Delta y, z_2) = \boldsymbol{FF}^{-1}[T(\Delta z)\boldsymbol{FF}[a(k\Delta x, l\Delta y, z_1)]]. \tag{5.14}$$

5.1 Sampling

There are two important and related issues in determining the numerical sampling. The highest spatial frequency which can be represented in the computer is determined by the sample spacing Δx and Δy. The region of space which can be represented is determined by the width of the computer array $M\Delta x$ and $N\Delta y$. First we will consider the diffraction phenomenology which can be observed with a given sample spacing.

The Nyquist sampling frequency — the highest frequency which can be represented — is

$$f_{\text{Nyquist}} = \frac{1}{2\Delta x}. \tag{5.15}$$

To understand how this function applies to diffraction patterns, consider the diffraction pattern of a top hat function. This function is both representative of many aperture functions and analytically tractable. Near the center of the pattern the characteristic frequency of the irradiance variations is $f = F_n/a$. This holds approximately true within the bright region of the geometrical beam. The frequency is $f = (2F_n)/a$, outside the geometrical beam — the shadow region. To satisfy the Nyquist frequency, the sampling should be chosen to be $\Delta x = a/2F_n$ to resolve the center detail, and $\Delta x = a/4F_n$ to resolve the detail in the shadow. These regions are illustrated in Fig. 4.7.

Failure to resolve the highest spatial frequencies may not result in an unacceptable representation of the function. In particular, using the near-field propagator for very short distances will show the distribution to be largely unchanged (the correct answer) even though the high spatial frequencies may not be correctly sampled.

Aliasing has a much more serious affect on the accuracy of the information. If the distribution grows outside the bounds of the array, severe aliasing will result which may render the calculation unusable. These errors arise from the finite size of the computer array. Because of propagation, a collimated beam expands and the complex amplitude grows beyond the bounds of the array and is folded back on itself, as shown in Fig. 5.4. This folded amplitude is the source of aliasing errors. The folded amplitude causes high spatial frequency errors in the intensity pattern, as shown in Fig. 5.4.

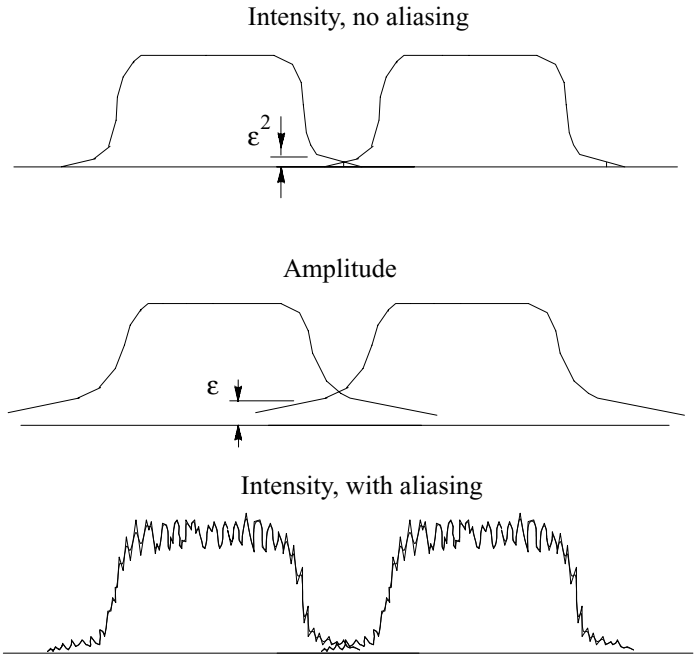


Fig. 5.4. Because of the periodic nature of the discrete calculations, the amplitude in the computer array folds over into neighboring “ghost” arrays and vice versa. The intensity distribution may be at a relatively small value, as indicated in the top drawing by ε^2. However, it is the complex amplitudes that interact at the boundary. The amplitudes decay much more slowly than the intensity, as indicated by ε in the middle drawing. An intensity pattern with aliasing is shown schematically in the bottom drawing. The aliasing errors are largest where the intensity is highest because the signal boosts the amplitude error.

The most severe errors tend to be where the distribution has the highest amplitude, not near the edge of the distribution. This is because the amplitude of the signal and error add rather than the intensities. Consider a nominally top hat function of unit amplitude and an aliasing contribution of ε. Assume that ε is slowly varying across the array. Near the edge of the array the intensity is of the order of magnitude of ε^2. In the center, the intensity is roughly $1 + 2\varepsilon + \varepsilon^2$. The error in the center will be of order 2ε — much larger than the error at the edge of order ε^2. It is important not to be deceived into believing that the aliasing errors are negligible by seeing an intensity distribution roll-off at the edge of the array. Approximate guidelines for the magnitude of aliasing errors may be determined for top hat functions, i.e., uniformly filled circular apertures. The results will be generally characteristic of distributions with strong discontinuities. For top hat functions there is an exact solution based on Lommel functions [3]. The Lommel functions may be approximated by an asymptotic solution [15]. This approximation enables the calculation of aliasing errors to be made for the bright region inside the geometric aperture area and the dark region in the shadow. Because the amplitude values add þ not the irradiance values — the aliasing errors are affected by the signal level. Let ε_b and ε_d be the errors in the bright and dark regions, then approximate expressions for the errors are

$$\varepsilon_b = 8\sqrt{\frac{1}{\pi^2 F_n}\frac{\frac{a^3}{r^3}}{1-\frac{a^2}{r^2}}}, \quad \varepsilon_d = \frac{3}{\pi^2 F_n}\frac{\frac{a^3}{r^3}}{1-\frac{a^2}{r^2}}. \tag{5.16}$$

These expressions give the order of magnitude of the effects. Aliasing errors are not always immediately distinguishable from diffraction ripples. Generally, high spatial frequency ripples will be manifest in the immediate vicinity of an aperture, but high spatial frequency aliasing errors will be present all over the distribution with the largest errors where the distribution has high intensity. Some experimentation with different size arrays and sampling may be required to gain an understanding of the appearance of the two phenomena. Consider a case of a beam of 5 mm diameter, wavelength 10.6 microns, and propagation of 5 cm. An array of 128×128 points is selected and in the first case, shown in Fig. 5.5.a, the array is almost completely filled by the aperture (78%).

In Fig. 5.5c the aperture is less than half the size of the array (39%). Both beams are propagated a distance of 10 cm, which is a Fresnel number of 5. Fig. 5.5c shows the diffracted beam with the small guardband. The distribution is filled with high frequency ripples. Fig shows the distribution with the large guardband but windowed to have the same size as Fig 5.5d. There is still a relatively large area surrounding the distribution of Fig 5.5c and the high frequency ripples are greatly reduced, but not completely absent.Examination of Figs. 5.5c and 5.5d shows that the aliasing errors are most noticeable in the center of the array and not at the edges as might be expected. This is because the aliasing errors add as complex amplitude numbers to the correct distribution and are effectively boasted by the distribution in the center of the array. It is clearly not sufficient to judge the degree of aliasing by the level of irradiance at the edge of the aperture.

Close examination of the degree of aliasing either by performing numerical experiments or using Eq. (5.16) may at first be discouraging since most near-field diffraction calculations have significant amounts of aliasing. In practice, many calculations are not adversely affected by significant levels of aliasing. For

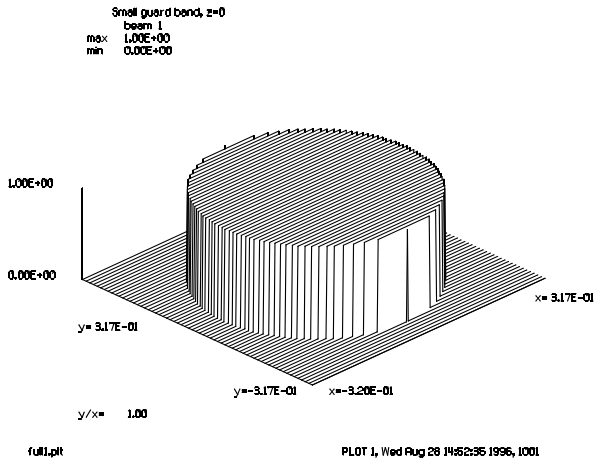


Fig. 5.5.a. The aperture is 0.5 cm in diameter and the width is 0.64 cm, leaving only a small guard band.

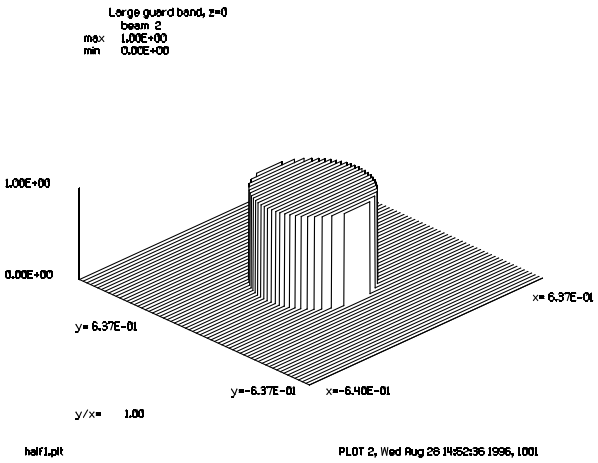


Fig. 5.5b. The aperture is the same size of.5 cm as Fig. 5.5c but the field width is 1.28 cm, so the guard band is much greater.

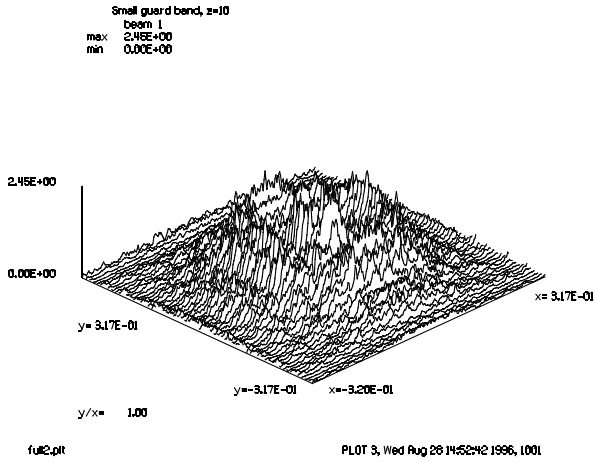


Fig. 5.5c. After propagating 10 cm (a Fresnel number of about 3.9) the distribution shows large scale diffraction ripples and many small aliasing ripples. Note that the ripples show up very strongly in the center and damp down toward the edge of the array.

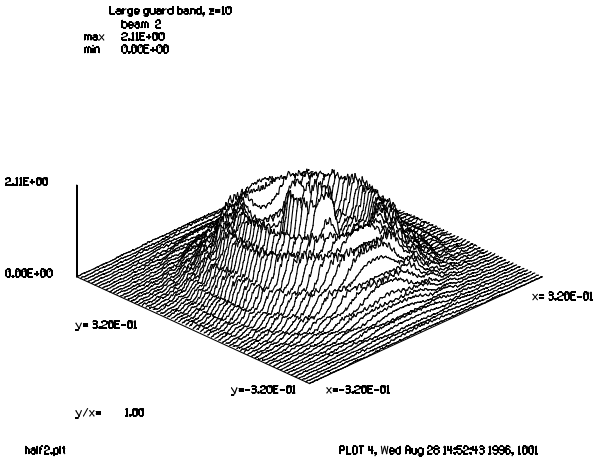


Fig. 5.5d. he beam with the large guard band, as shown in Fig. 5.5c. has been propagated the same distance. The plot shows the distribution rescaled to cover the same size as Fig. 5.5c. for direct comparison. There still remains a large guard band around the distribution which is essentially empty.

specific problems, one can try various guardband values to determine whether results are affected. Ideally one increases the array size and the guardband width þ keeping the same number of sample points across the distribution until no appreciable change in the results is observed. In practice one may choose the array size based on the computational time that is consistent with one's own level of patience.

5.2 Automatic Selection of Matrix Units

This section describes an automatic selection algorithm for sampling units that is made in choosing a gaussian beam. Since the authors do not know in advance what the user is going to do with the beam distribution once it is formed, the only reasonable choice for an automatic program choice of units is to make

the sampling the same in both the spatial and frequency domains. To do so, we must make some assumptions about the beam distribution. Consider a gaussian beam and its Fourier transform. Further assume $\Delta x = \Delta y = \Delta r$,

$$a(k\Delta r) = e^{-\left(\frac{k\Delta r}{\omega_0}\right)^2}, \tag{5.17}$$

$$a(m\Delta\rho) = \pi\omega_0^2 e^{-\left(\pi\omega_0\frac{m}{N\Delta r}\right)^2}, \tag{5.18}$$

ω_0 = gaussian beam waist,
k = spatial domain index,
m = frequency domain index.

For equal sampling in both domains, the functions $a(r)$ and $a(\rho)$ would have the same relative value when $k = m$. The solution for Δr to accomplish this is

$$\Delta r = \omega_0\sqrt{\frac{\pi}{N}}. \tag{5.19}$$

A similar solution for a top hat function can be derived as well. Given a top hat of radius a_0 and solving for the condition that the top hat function and the first Airy dark ring should span the same number of sample points, we have

$$\Delta r \approx a_0\sqrt{1/61N}. \tag{5.20}$$

Top hat and gaussian functions with equal sampling are illustrated in Figure 5.6.

5.3 Propagation Control

The previous section establishes an acceptable initial sampling condition. However, the beam spreads due to diffraction and may therefore overfill the computer array. Fortunately the near-field and far-field propagators may be used to control the size of the array so that the beam aliasing does not change much from the initial state. The sampling period of the near-field is constant. The sampling period of the far-field is

$$\Delta x_2 = \frac{\lambda|\Delta z|}{M\Delta x}. \tag{5.21}$$

By use of a combination of near-field and far-field propagators, the sampling period may be set to any required value. It is necessary to have some analytical measure of the size of the diffracted beam at all points in space, so that the correct sampling period may be chosen. The Fresnel number may seem to be an obvious choice. In conventional optical systems the beam stays in the near-field (large Fresnel numbers) as it goes through all the optical elements and then propagates to the far-field (small Fresnel numbers) after the final

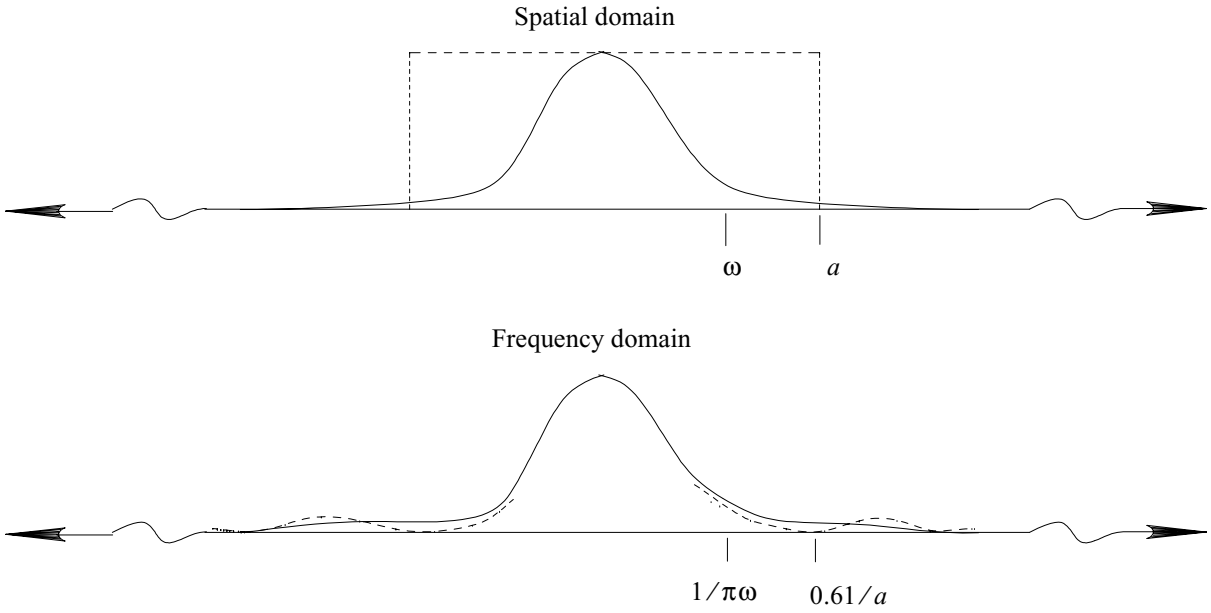


Fig. 5.6. This figure shows schematically how a gaussian (solid lines) and a top hat function (dashed lines) and their Fourier transforms appear if the domains have equal sampling.

aperture. More complex systems may have internal foci so that the beam goes from near- to far-field inside the beam train. Some systems may be analyzed accurately ignoring all apertures. In that case, it is not possible to define near- and far-field regions and the concept of Fresnel number is meaningless. The conditions were illustrated in Fig. 4.4. The gaussian beam has an easily calculated width at all points in space. Any complex amplitude distribution in space may be approximately fit to a gaussian beam of radius, ω, and phase radius, R. From these value, the gaussian waist size ω and the distance to the waist z may be calculated by Eq. (4.27). The gaussian beam acts as a surrogate to the actual beam. The gaussian beam parameters should be recalculated, so that the surrogate beam remains representative of the actual distribution.

We could elect to adjust the sampling period to track exactly. In this case the units would be continuously changing. It has proved to be more desirable to have a region of constant sampling period near the waist and a region of linearly increasing sampling period far from the waist. The two possibilities are illustrated in Fig. 5.7. The far-field propagation has an expanding coordinate system of the form of Eq. 5.21 and the near-field propagation has a constant coordinates system of the form, $\Delta x_2 = \Delta x_1$.

When the beam size becomes comparable to the array size, aliasing occurs. To calculate the distance at which the switch should be made from constant sampling period to linearly increasing sampling period, we solve for the distance from the waist at which the increasing period equals the constant period. If $\Delta x_v = \lambda z / M \Delta x$ for the variable sampling period, then $\Delta x_v = \Delta x$ is the condition to be met for change from the constant to variable sampling periods. This occurs when $z = z_R$. The switch is made at the Rayleigh range. This condition is illustrated in Figure 5.8. The process can, of course, be reversed. We can propagate from any distant point back to the waist.

A function $f(x, y)$ may be defined as the complex amplitude with respect to the curved reference surface of radius z such that

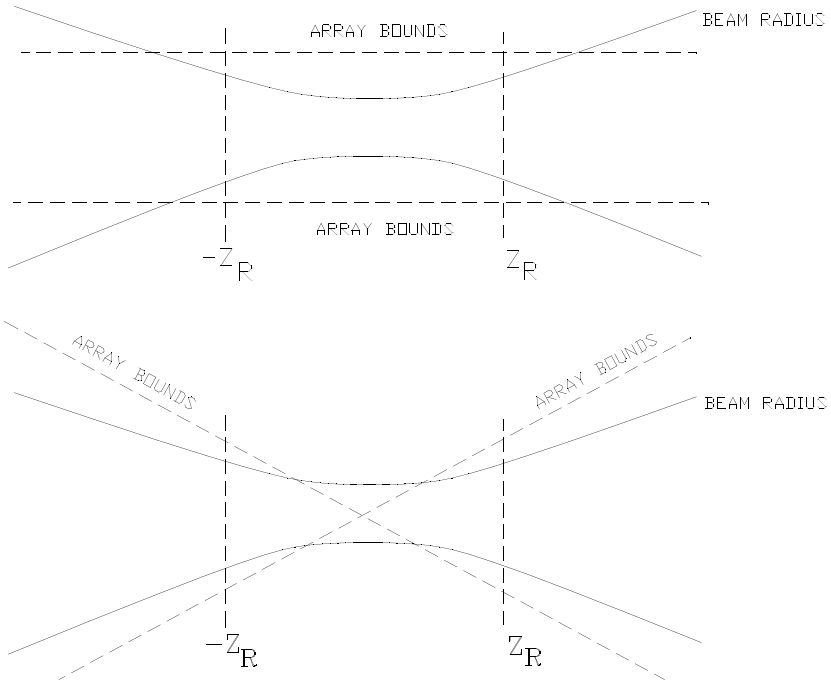


Fig. 5.7. Dashed lines represent fixed (upper) and variable (lower) array sizes.

$$f(x,y) = a(x,y)e^{j\frac{kr^2}{2z}}, \tag{5.22}$$

where z is still the distance from the waist. We can propagate either $a(x,y)$, referenced to a plane surface, or $f(x,y)$, referenced to the curved surface, using the equations to be presented. We wish to select either $a(x,y)$ or $f(x,y)$ depending on which has the smaller residual phase. The surrogate gaussian beam again proves to be useful. At any point in space, the gaussian beam is

$$a(r) = e^{-\frac{r^2}{\omega^2}} e^{-j\frac{kr^2}{2R}}. \tag{5.23}$$

The function $f(r)$, using the curved reference is

$$f(r) = e^{-\frac{r^2}{\omega^2}} e^{-j\frac{kr^2}{2R}} e^{j\frac{kr^2}{2z}}. \tag{5.24}$$

The critical question is whether the residual phase of Eqs. (5.23) or (5.24) is less. Consider the phase error of the actual wavefront with respect to either a planar or spherical reference surface, evaluated at the $1/e$ point of the gaussian amplitude,

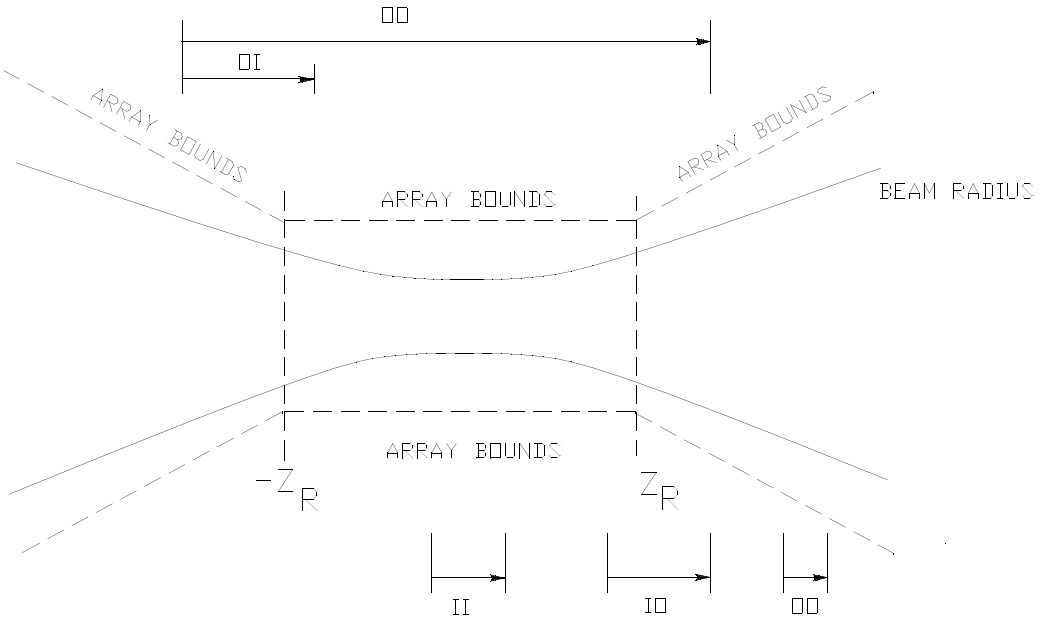


Fig. 5.8. Array size with constant size inside the Rayleigh range and variable size outside. There are four different possibilities in moving from inside or outside to inside or outside indicated by the acronyms II, IO, OI, OO.

$$\Delta W_{\text{plane}} = -\frac{h^2}{2}\frac{1}{R}\bigg|_{h=\omega(z)}, \quad \Delta W_{\text{sphere}} = -\frac{h^2}{2}\left(\frac{1}{R}-\frac{1}{z}\right)\bigg|_{h=\omega(z)}, \tag{5.25}$$

$$\Delta W_{\text{plane}} = -\frac{\omega_0^2}{2}\frac{z}{z_R^2}, \quad \Delta W_{\text{sphere}} = \frac{\omega_0^2}{2}\frac{1}{z}. \tag{5.26}$$

These functions are plotted in Fig. 5.9. The phase error ΔW is minimized by choosing a plane reference inside the Rayleigh distance and a spherical reference outside the Rayleigh distance.

For a gaussian beam, $a(r)$ should be used if $|z| \le z_R$. For $|z| \ge z_R$, the $f(r)$ form should be used. This quadratic phase factor is exactly the same as is applied in the far-field propagator Eq. (5.13a), so that phase factor need not be applied.

The objective of the propagation control algorithm is to be able to move from any point in space to any other. To do so it is convenient to first define primitive operators; plane-to-plane (PTP), waist-to-spherical (WTS), and spherical-to-waist (STW).

Let us first find the best fit gaussian function which represents the current state of the beam width and phase radius. Let the best fit gaussian parameters be ω_1 and R_1. Find the distance to the gaussian waist (in the current optical space) and the gaussian waist radius by using Eq. (4.32)

We shall consider whether a given point lies inside or outside the Rayleigh range

$$\text{inside } |z - z_\omega| \le z_R, \tag{5.27}$$

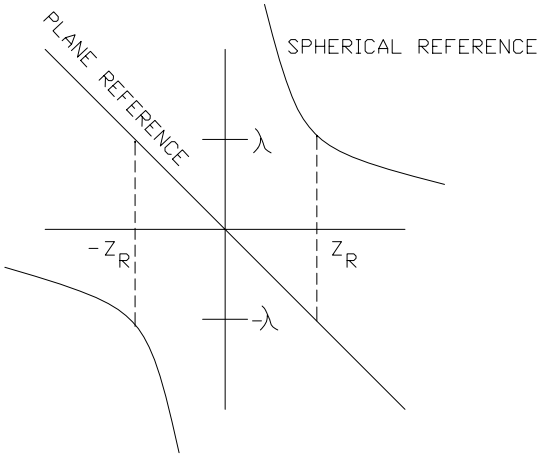


Fig. 5.9. Phase at the 1/e radius point as a function of distance from the waist. The plane reference surface has a small phase error near the waist. A spherical reference surface has the smallest phase error for large distances from the waist. At the Rayleigh distance, the two reference surfaces have phase errors of equal magnitude but opposite sign.

$$\text{outside } |z - z_R| \geq z_R \,. \tag{5.28}$$

We shall now classify the propagation into whether the starting and ending points are inside or outside. We define four operators which cover all possible cases are

$$\mathbf{II}(z_1, z_2) : \text{inside } z_R \text{ to inside } z_R \,, \tag{5.29}$$

$$\mathbf{IO}(z_1, z_2) : \text{inside } z_R \text{ to outside } z_R \,, \tag{5.30}$$

$$\mathbf{OI}(z_1, z_2) : \text{outside } z_R \text{ to inside } z_R \,, \tag{5.31}$$

$$\mathbf{OO}(z_1, z_2) : \text{outside } z_R \text{ to outside } z_R \,. \tag{5.32}$$

These operations are illustrated in Fig. 5.8. We shall define these four operators in terms of three primitive operators:

$$\mathbf{II}(z_1, z_2) \;=\; \mathbf{PTP}(z_2 - z_1) \,, \tag{5.33}$$

$$\mathbf{IO}(z_1, z_2) \;=\; \mathbf{WTS}(z_2 - z_\omega)\mathbf{PTP}(z_\omega - z_1) \,, \tag{5.34}$$

$$\mathbf{OI}(z_1, z_2) \;=\; \mathbf{PTP}(z_2 - z_\omega)\mathbf{STW}(z_\omega - z_1) \,, \tag{5.35}$$

$$\mathbf{OO}(z_1, z_2) \;=\; \mathbf{WTS}(z_2 - z_\omega)\mathbf{STW}(z_\omega - z_1) \,. \tag{5.36}$$

These primitive operators are

$$\mathbf{PTP}(\Delta z) = \mathbf{FF}^{-1}[T(\Delta z)\mathbf{FF}[(\ )]],\ \Delta x_2 = \Delta x_1 .\ \Delta y_2 = \Delta y_1, \tag{5.36a}$$

$$T(\Delta z) = e^{-j\pi\lambda\Delta z\rho^2}, \tag{5.36b}$$

$$\mathbf{WTS}(\Delta z) = \frac{1}{j\lambda\Delta z}\mathbf{FF}^{s}[q(\Delta z)(\ )],\ \Delta x_2 = \frac{\lambda|z|}{M\Delta x_1},\ \Delta y_2 = \frac{\lambda|z|}{N\Delta y_1}, \tag{5.37a}$$

$$q(\Delta z) = e^{j\frac{\pi r^2}{\lambda\Delta z}}, \tag{5.37b}$$

$$s = \frac{\Delta z}{|\Delta z|}, \tag{5.37c}$$

$$\mathbf{STW}(\Delta z) = \frac{1}{j\lambda\Delta z}q(\Delta z)\mathbf{FF}^{s}[(\ )],\ \Delta x_2 = \frac{\lambda|z|}{M\Delta x_1},\ \Delta y_2 = \frac{\lambda|z|}{N\Delta y_1}. \tag{5.38}$$

These can also be written in separable form. The phase bias operations associated with a lens and incident light which may or may not have a radius of curvature and for exiting light which may or may not have a radius of curvature are illustrated in Fig. 5.10. The operators for the four possible cases are described in Table 5.1

Table. 5.1. Phase corrections applied to the array after a lens depending on surrogate gaussian parameter.

inside-to-inside	$t(x,y) = e^{-j\frac{kr^2}{2}\frac{1}{f}}$	no phase bias
inside-to-outside	$t(x,y) = e^{-j\frac{kr^2}{2}\left(\frac{1}{f}+\frac{1}{R'}\right)}$	phase bias after lens
outside-to-inside	$t(x,y) = e^{-j\frac{kr^2}{2}\left(\frac{1}{f}-\frac{1}{R'}\right)}$	phase bias before lens
outside-to-outside	$t(x,y) = e^{-j\frac{kr^2}{2}\left(\frac{1}{f}-\frac{1}{R}+\frac{1}{R'}\right)}$	phase bias before and after lens

5.3.1 Using Surrogate Gaussian Beams for Propagation Control

Surrogate gaussian beams are used for propagation control. These beams are considered to behave approximately as the actual beam behaves and, therefore, allow a convenient method of determining the algorithms to be used. There is some latitude in the selection of the algorithms for most problems and the algorithm choice need not be determined from calculations of great accuracy. The exception to this is

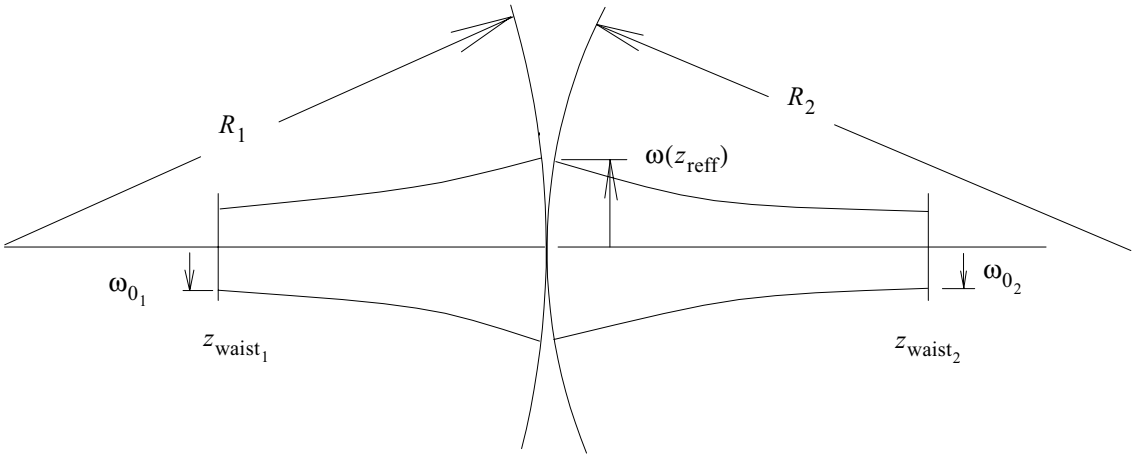


Fig. 5.10. Relationship between surrogate beam parameters before and after a lens.

resonators where care should be taken to use exactly the same algorithm choices in each pass to achieve maximum numerical stability. This may be done with the `zbound` or `geodata` commands.

When the beam is initialized by the `clap` or `gaussian` commands a choice is made of the surrogate gaussian size.Subsequent propagation distances and optical elements will modify that choice. One may update the surrogate gaussian beam parameters fitting the actual information in the beam array with `fitgeo` or explicitly by using `geodata`. The defining parameters of the surrogate gaussian beam are the waist ω_0 and the position of the waist z_ω. The surrogate gaussian waist radius and position do not change with propagation. An optical element will change the phase radius of curvature at the lens by the Lens Law. The change to the surrogate beam by the lens takes place in five steps:

1) Calculate incident phase radius of curvature R_1 and local transverse radius according to Eq. (4.31)

$$R_1 = (z_{\text{reff}} - z_\omega) + \frac{z_R^2}{z_{\text{reff}} - z_\omega}, \quad \omega(z_{\text{reff}}) = \omega_0 \sqrt{1 + \left(\frac{\lambda(z_{\text{reff}} - z_\omega)}{\pi\omega_0^2}\right)^2}, \tag{5.39}$$

2) Calculate new radius of curvature R_2 by the Lens Law

$$\frac{1}{R_1} - \frac{1}{R_2} = \frac{1}{f}, \tag{5.40}$$

3) Calculate the new surrogate gaussian parameters z_{ω_2} and ω_{0_2} from Eq. (4.32):

$$z_{\omega_2} - z_{\text{reff}} = -\frac{R_2}{1 + \left(\frac{\lambda R_2}{\pi\omega z_{\text{reff}}^2}\right)^2}, \quad \omega_0 = \frac{\omega z_{\text{reff}}^2}{\sqrt{1 + \left(\frac{\pi\omega z_{\text{reff}}^2}{\pi R_2}\right)^2}}, \quad z_{R_2} = \frac{\pi(\omega_{0_2})^2}{\lambda}, \tag{5.41}$$

4) Determine whether new beam is inside or outside new Rayleigh distance

$$\text{inside if } \left|z_\omega^{(+)} - z_{\text{reff}}\right| < z_R^{(+)}, \text{ outside if } \left|z_\omega^{(+)} - z_{\text{reff}}\right| > z_R^{(+)}, \tag{5.42}$$

5) Apply phase correction factors to the distribution depending on the old inside-outside and new inside-outside status.

The set of steps described above must be done for each element with optical power. Let us consider a simple example. Consider the gaussian beam of Example 3. Given a gaussian beam of wavelength 10.0μ, initial waist $\omega_0 = 0.5$, located at $z_\omega = 0$, propagation of 100 cm, lens of $f = 100$ cm and propagation of 100 cm. This results in a waist at the last plane. The surrogate gaussian beam proceeds in a slightly different fashion as indicated in Table 5.2. In this example the lens lies inside the initial Rayleigh range and outside the final Rayleigh range so a quadratic phase factor, according to Table 5.2, is applied.

Table. 5.2. Example of surrogate gaussian beam propagation, from Example 3.

Event	**IO status**	z_R	ω_0	z_ω	ω	R
start	I	785.398	0.50000	0.0	0.50000	∞
prop, 100	I	785.398	0.50000	0.0	0.50404	6268.505
lens, $f = 100$	O	12.732	0.063662	200.	0.50404	-101.6211
prop, 100	O	12.732	0.063662	200	0.063662	∞

5.4 Finite Difference Propagation

High Fresnel diffraction and high spatial frequencies require high sampling rates and, necessarily, large arrays. FFT computations are time consuming and the question naturally arises as to whether alternate methods have advantages. The method most frequently considered is the finite difference propagator (FDP). The advantages of the FDP are that it is very fast for short steps [2]. The disadvantages are

1) one must take short steps (repeated application of the algorithm is required for long propagation),
2) the algorithm is numerically unstable at discontinuities,
3) certain diffraction effects are washed out.

This length is called the characteristic diffraction length. Since the iterative solution of the propagation problem in a nonlinear active medium requires recalculation of the diffraction effects every characteristic length, it may be advantageous to use the FDP in solution of the inhomogeneous wave equation. Because of the restrictions on step length and discontinuities, considerable care must be exercised in the use of the FDP.

5.4.1 Real-Imaginary Description

The finite difference propagator was developed by Rench[2]. The derivation is very simple. Rather than use the plane wave decomposition, we solve the parabolic wave equation directly:

$$\frac{\partial A}{\partial z} = \frac{1}{2jk}\left(\frac{\partial^2}{\partial^2 x} + \frac{\partial^2}{\partial^2 y}\right)A . \tag{5.43}$$

The second derivatives are taken from the discrete complex amplitude distribution by

$$\frac{\partial^2 A}{\partial x^2} \approx \frac{A(I+1,J) + A(I-1,J) - 2A(I,J)}{2\Delta x}, \tag{5.44}$$

$$\frac{\partial^2 A}{\partial y^2} \approx \frac{A(I,J+1)1 + A(I,J-1) - 2A(I,J)}{2\Delta y}, \tag{5.45}$$

where $A(I,J)$ is an element of the complex amplitude array and I and J are the indices. The second derivative is taken by considering only the neighboring points. In the area of a discontinuity the numerical procedure blows up. One can patch the algorithm somewhat but there are no perfect fixes known. The major limitation is, therefore, errors which occur at sharp boundaries, such as the edge of mirrors. The principal manifestation of the intensity discontinuity is a large phase spike. This phase spike will decay during subsequent propagation steps but will spread sideways, ultimately causing serious disruption of the distribution. The width of the damaged region increases each application of the algorithm. A second problem, unrelated to the discontinuity problem arises from the fact that the finite difference propagator must be applied several times per characteristic diffraction length. The longest single allowable propagation step was calculated by Rench to be

$$\Delta z < \frac{k}{2}\Delta x^2 . \tag{5.46}$$

Taking $2\Delta x = 1/f$, we find Eq. (5.42) very similar to Eq. (4.14). To propagate any significant distance the algorithm must be repeated many times.

For strong nonlinear gain, we may have to calculate diffraction and gain intermittently at many points along the axis, taking steps no longer than the characteristic length. In that case, the requirement for short steps with the finite difference propagator is not a problem. Figures 5.11.a and 5.11b show propagation of the same distance with the FFT and the finite difference algorithms respectively. The finite difference propagator can certainly be improved by using good numerical analysis methods but the method will eventually manifest numerical instabilities after a sufficiently long propagation sequence.

5.4.2 Amplitude-Wavefront Description

The amplitude-wavefront representation provides an alternate method of representing the optical beam: complex amplitude $A\exp(jW)$, where A is the amplitude and W is the wavefront error. In amplitude-wavefront, Eq. (5.39) takes the form

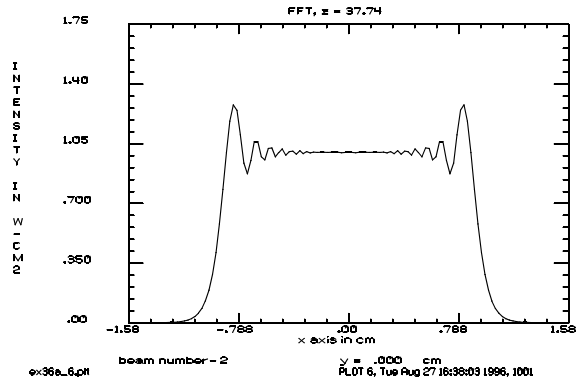


Fig. 5.11.a. Propagation with the FFT algorithms.

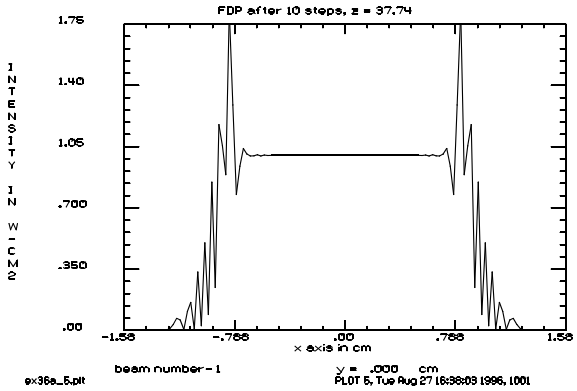


Fig. 5.11b. Propagation with a finite difference propagator the same distance as Fig. 5.11.a. The spurious errors near the edge are due to the algorithm.

$$\frac{\partial A}{\partial z} = \frac{A}{2}\left(\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2}\right) + \frac{\partial A}{\partial x}\frac{\partial W}{\partial y} + \frac{\partial A}{\partial y}\frac{\partial W}{\partial x} \tag{5.47}$$

$$\frac{\partial W}{\partial z} = \frac{1}{2}\left[\left(\frac{\partial W}{\partial x}\right)^2 + \left(\frac{\partial W}{\partial y}\right)^2\right] - \frac{1}{2Ak^2}\left(\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2}\right) \tag{5.48}$$

Equations (5.43) and (5.44) allow direct propagation of the amplitude-wavefront representation of the beam. It is capable of representing large wavefront errors without the problem of higher order intrinsic to complex amplitude form. It is, however, subject to the same difficulties of finite difference propagation as Eq. (5.43).

5.5 Propagation of Rotational Functions and Fast Hankel Transforms

The primary propagation algorithms in GLAD are based on rectangular propagation with angular spectrum techniques that assume no symmetry. In the case of strictly rotationally symmetric functions, Hankel transforms may in principle be used to solve the diffraction integrals. The Hankel transform pair may be written:

$$A(\rho) = 2\pi\int_0^\infty a(r')J_0(2\pi\rho r')r'\mathrm{d}r', \tag{5.49}$$

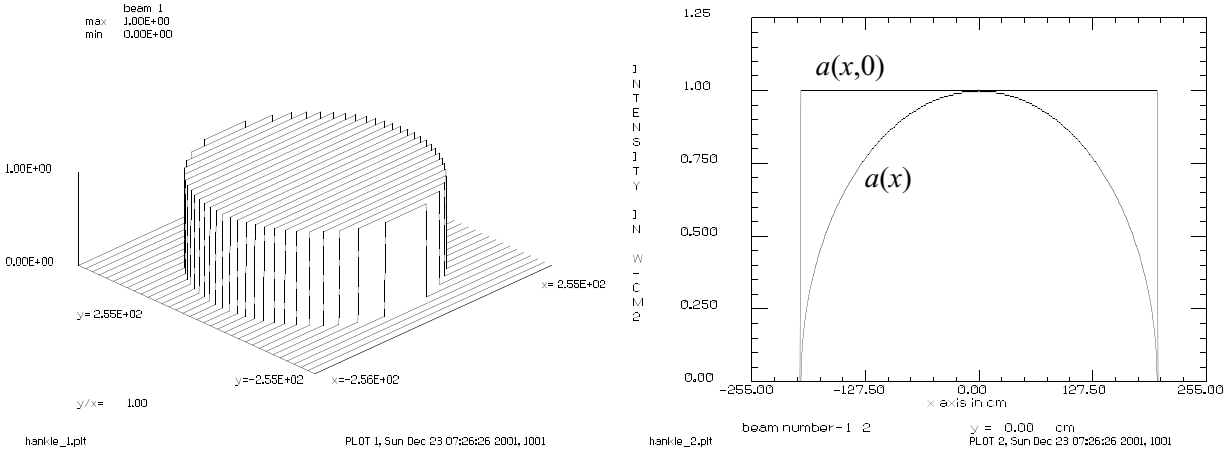
$$a(\rho) = 2\pi\int_0^\infty A(r')J_0(2\pi r\rho')\rho'\mathrm{d}\rho'. \tag{5.50}$$

Direct solution in terms of the Bessel function representation is relatively slow. The fast-Hankel transform was devised by Siegman to provide a faster method[4–7]. This method has proved successful in many cases but suffers in numerical implementation from a singularity at zero radius in both spatial and

frequency domains and nonlinear sample spacing. The zero radius singularity may be minimized by good programming but still has results in some mean power uncertainty. This often manifests itself as “edge droop” or “slopping shoulders” in the diffraction pattern, even for very short propagation steps. The zero radius singularity makes precise energy conservation difficult. The nonlinear sampling results in higher sampling densities at the edge of the array which is an advantage in some cases but makes it awkward to interface with the uniformly sampled points in rectangular array codes such as GLAD. A highly efficient circular propagator with uniform sample spacing and no zero radius singularity resulting in excellent energy conservation. The method is based on a degenerate form of two-dimensional Fourier transform. In the general case, a two-dimensional Fourier transform is used to calculate a two dimensional frequency spectrum of form $A(\xi, \eta)$. If, however, we only need the frequency spectrum along a single row where $\xi = 0$, we can simplify the two-dimensional Fourier transform into a sum along the y-direction and a one-dimensional transform

$$A(\xi, 0) = \iint a(x, y) e^{-j2\pi(x\xi + y\eta)} \mathrm{d}x\mathrm{d}y \Big|_{\xi = 0} = \int a(x) e^{-j2\pi x\xi} \mathrm{d}x \,. \tag{5.51}$$

where $a(x) = \int a(x)\mathrm{d}y$. $a(x)$ is the sum along the y-direction. The y-sum $a(x)$ can be quickly computed from the center row of the square array $a(x, 0)$ by interpolation to find the values at the various (x, y) points. The projection method based on y-sums may be used to calculate a Hankel transform pair between $a(x, 0)$ and $A(\xi, 0)$. Figure Fig. 5.12a illustrates a two-dimensional function $a(x,y)$ in the form of a circular distribution and Fig. 5.12b shows a scan through the circular distribution, $a(x,0)$ and the y-integrated function $a(x)$. For an array size of 1024 x 1024 an improvement in speed of between 20 and 40 times may be realized doing the one-dimensional transformation of $a(x)$ rather than the two-dimensional transforms of $a(x,y)$. Example 96 illustrates the propagation of circular arrays and compares the accuracy with full two-dimensional calculations. Fig. 5.13 compares calculations with square and degenerate FFT methods.



(a) (b)

Fig. 5.12. A circular distribution (a) is represented by a diagonal scan $a(x,0)$ in (b). The y-intregated distribution of $a(x,0)$ is $a(x) = \int a(x)\mathrm{d}y$ as illustrated by the curved line. One dimensional Fourier transformation of a(x) yields a scan through the Frurier transform of $a(x,y)$.

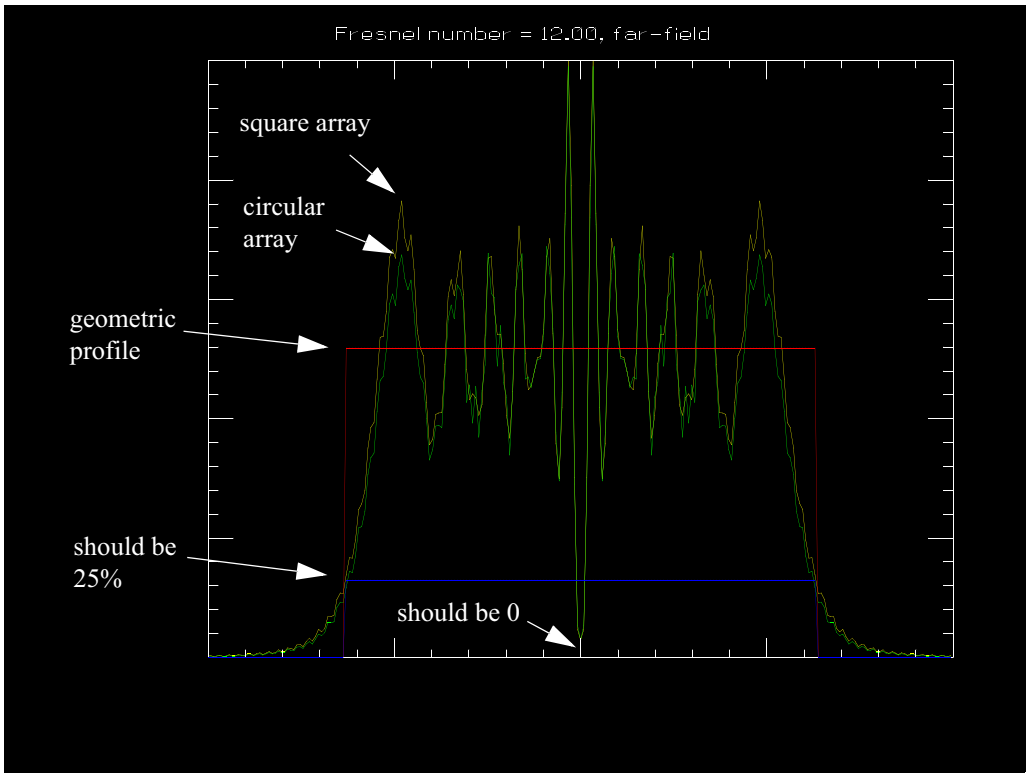


Fig. 5.13. Illustration of Fresnel number 12 in a 1024 ×1024 array (yellow) and the method of degenerate FFT (green) using and a 1024 ×1 array. Some edge droop is evident with the degenerate FFT method due to interpolation.

References

1. J. Hayes, “Fast Fourier Transforms and their Applications,” Applied Optics and Optical Engineering, Vol. 11, Eds. R. Shannon and J. Wyant, Academic Press (1991).
2. D. B. Rench and Chester, “Three-dimensional unstable resonators with laser medium,” Appl. Opt. 13, p2546-2561 (1974).
3. M. Born and E. Wolf, *Principles of Optics*, Pergammon Press, New York 1965
4. A.E. Siegman, “Quasi fast Hankel transform,” Optics Lett. 1, 13–15 (March, 1977).
5. Shinan-Chur Sheng, Chap. 3, “Studies of Laser Resonators and Beam Propagation Using Fast Transfrom Methods,” PhD Dissertation, Dept. of Applied Physics, Stanford Univ., March 1980.
6. S-C. Sheng and A. E. Siegman, “Nonlinear optical calculations using fast transform methods: Second harmonic generation with depletion and diffraction,” Phys. Rev. A 21, 599-606 (February 1980).
7. A. V. Oppenheim, G. V. Frisk, and D. R. Martinez, “Computation of the Hankel transform using projections,” J. Acoust. Soc. Am. 68, 523-529 (1980).

6. Optical Systems and Inhomogeneous Media

Geometrical analysis often considers an optical system in terms of its ideal imaging behavior, describable by paraxial optics and referred to as the stigmatic properties, and the aberrations or non-stigmatic behavior. One of the definitions of an aberration-free optical system is that its geometrical imaging properties, as determined by aberration polynomial calculations or the tracing of real rays, agree with the paraxial behavior. In physical optics propagation, we can proceed in similar fashion and realize a great simplification of many problems. Often, many optical elements may be combined into a single equivalent operator and gradient index optics may be treated with the same facility as conventional components.

6.1 ABCD Operators

The matrix method, sometimes called the ABCD matrix method, is a convenient method of defining the paraxial behavior of an optical system. Some common matrix operators are listed in Table 6.1 for reference including translation, refraction at a spherical surface, reflection at a spherical surface, refraction in a quadratic inhomogeneous media, and refraction in the Maxwell's fisheye lens. The ABCD matrix methods aptly describes the paraxial behavior, the gaussian beam properties, and the general diffraction behavior [1, 3]. By couching the diffraction calculations in terms of geometrical paraxial optics; the calculations can be simplified and it becomes clear that strongly inhomogeneous media can be as readily treated in the same fashion as conventional elements. Near-stigmatic systems include, free-space, homogeneous media, systems of lenses and mirrors, and gradient refractive index (GRIN) optics of all well-behaved types.

The ABCD operators may be written for transformation of ray heights and angles

$$\text{ray heights and angles, } \boldsymbol{r} = \begin{pmatrix} y \\ u \end{pmatrix} \tag{6.1}$$

in the manner of Yariv[4] or in terms of reduced lengths and reduced angles

$$\text{reduced length and optical angles, } \tau = \begin{pmatrix} y \\ nu \end{pmatrix} = \begin{pmatrix} y \\ \omega \end{pmatrix} \tag{6.2}$$

in the manner of Siegman [1]:

$$\text{Yariv } \begin{pmatrix} y' \\ u' \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} y \\ u \end{pmatrix} \text{ , and Siegman } \begin{pmatrix} y' \\ n'u' \end{pmatrix} = \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} \begin{pmatrix} y \\ nu \end{pmatrix}, \tag{6.3}$$

where $A = A'$, $B = B'n'$, $C = C'/n'$, and $D = D'n/n'$. In GLAD, Yariv's method is used to show the true value of the angles explicitly.

An equivalent, elementary optical system can be found for any ABCD matrix. Let the properties of the first optical space be indicated by the subscript 1 and the properties of the second space by the subscript 2.

It is assumed that the ABCD matrix consists of four elementary operations: magnification change, change of index, thin lens of optical power ϕ, and a translation of distance t

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\varphi/n_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & n_1/n_2 \end{pmatrix} \begin{pmatrix} M & 0 \\ 0 & 1/M \end{pmatrix}. \tag{6.4}$$

One could choose a different set of four primitive operations, for example, by permuting the order of the matrices, and develop the method in similar fashion to that done here. The elementary operations are illustrated in Table. 6.1 and illustrated in Fig. 6.1. Siegman describes a generalized diffraction operator using these ABCD matrices explicitly and alternate approach, [1]. .

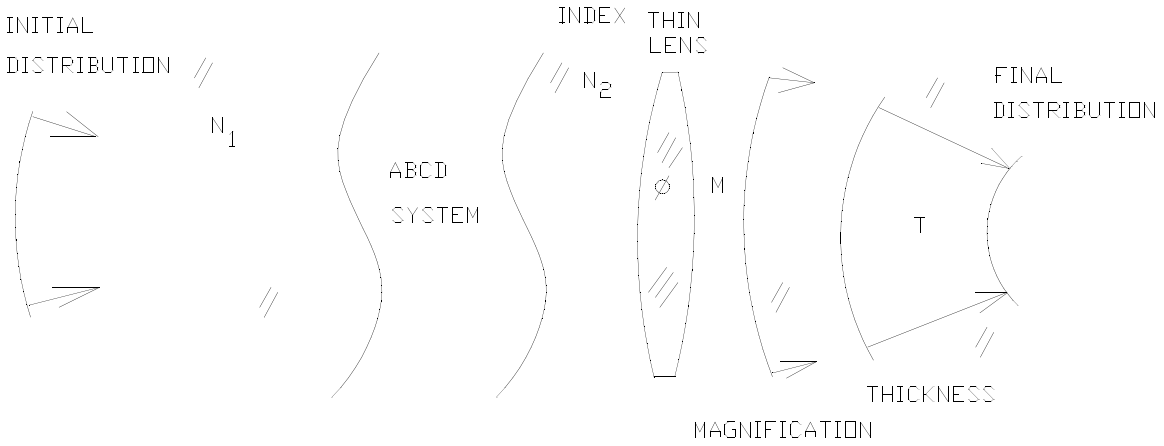


Fig. 6.1. .The paraxial behavior of an arbitrary optical system may be characterized by an ABCD matrix. Any ABCD matrix may be factored into four elementary operations change in index change, magnification, and divergence and a propagation step.

$$\tilde{u}_2(x_2) = \sqrt{\frac{j}{B\lambda_0}} \int_{-a_2}^{a_1} \tilde{u}_1(x_1) \exp\left(-j\frac{\pi}{B\lambda_0}(Ax_1^2 - 2x_1x_2 + Dx_2^2)\right) \mathrm{d}x_1 . \tag{6.5}$$

Siegman also considers implementation of phase bias so that the method is comparable to the **OO** operator or the **II** operators, described in Section 3, when no phase bias is used. The method does not lend itself to crossing the Rayleigh range boundaries and, therefore, is less general.

Equation (6.4) may be interpreted as finding the image of the original distribution in the local optical space, and completing the propagation to the desired plane. Provided the apertures in the intervening system are oversized, there are no diffraction propagation effects in transforming the object space distribution to the local image — only the magnification and radius change. All diffraction propagation effects will occur in the single propagation step of distance t using the diffraction propagation theory developed in the previous section.

Table. 6.1. Some ABCD matrix operators

ray vector, $\begin{pmatrix} y \\ u \end{pmatrix}$, y is the ray height, u is the angle

translation, $\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$, t is the translation distance

thin lens in air, $\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\phi & 1 \end{pmatrix}$, thin lens embedded in medium of index n,

$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\phi/n & 1 \end{pmatrix}$

zero power refractive boundary, $\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & n/n' \end{pmatrix}$, powered refractive boundary,

$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{\phi}{n'} & \frac{n}{n'} \end{pmatrix}$

n and n' are the indices before and after refraction, ϕ is the optical power, and $\phi = \frac{n'-n}{R}$

lens-like media, $n(r) = n_0 - \frac{1}{2}\gamma^2 r^2$, $\gamma^2 = \frac{2\Delta n}{r_0^2}$, $\mathrm{ABCD}(z) = \begin{pmatrix} \cos(\gamma z) & \sin(\gamma z)/\gamma \\ -\sin(\gamma z)/\gamma & \cos(\gamma z) \end{pmatrix}$

Maxwell's fisheye, $n(r) = \frac{n_0}{1+\frac{r^2}{f^2}}$, $\mathrm{ABCD}(z) = \frac{1}{z_1^2+f^2}\begin{pmatrix} 2z_1z_2 - z_2^2 + f^2 & z_1z_2^2 - z_2z_1^2 + (z_2-z_1)f^2 \\ 2(z_1-z_2)f^2 & 2z_1z_2 - z_1^2 + f^2 \end{pmatrix}$

The solution for the four free variables of Eq. (6.4) is

$$\frac{n_1}{n_2} = AD - BC, \tag{6.6}$$

$$M = \frac{AD-BC}{D}, \tag{6.7}$$

$$\frac{\phi}{n_2} = -\frac{CD}{AD-BC} = -\frac{C}{M}, \tag{6.8}$$

$$t = \frac{B}{D}. \tag{6.9}$$

The first three steps are quite simple to execute. The last step is a propagation of distance i in general, noncollimated geometry, as described in Chapter 5. Only this step requires any substantial computation time.

In the case of propagation to an exact focus point $D \to 0$ and the previous treatment is ill-defined because of the division by D

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\varphi_2/n_2 & 1 \end{pmatrix}\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -\varphi_1/n_2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & n_1/n_2 \end{pmatrix}, \tag{6.10}$$

$$\frac{n_1}{n_2} = AD-BC, \tag{6.11}$$

$$\frac{\phi_1}{n_2} = \frac{(1-A)}{B}(AD-BC), \tag{6.12}$$

$$t = \frac{B}{(AD-BC)}, \tag{6.13}$$

$$\frac{\phi_2}{n_2} = \frac{AD-BC-D}{B}. \tag{6.14}$$

Mirrors, lenses, and refractive surfaces in the paraxial approximation can be modeled using the focal length of the component. Clearly not all components are aberration free. The aberration may be explicitly added provided it is known by measurement or calculated analytically or by a ray tracing code. The paraxial phase effect is

$$t(x,y) = e^{-j\frac{k}{2f}(x^2+y^2)}, \tag{6.15}$$

where $t(x,y)$ is the complex transmission function, k is the wave number and, f is the mirror radius.

The focal length is defined with respect to the element's effect on the beam

$f > 0$ causes the light to be more convergent,

$f < 0$ causes the light to be more divergent.

For either lenses or mirrors the phase imposed by the element may be applied as a change to the complex amplitude distribution or as a change to the phase bias, as described by

Eqs. (5.22) to (5.24). The proper choice is to select the phase bias to achieve the minimum phase curvature that exists in the complex amplitude to minimize sampling requirements. To determine the proper choice of sampling condition after the optical element, we calculate the radius of curvature in image space by the equation,

$$\frac{1}{R} - \frac{1}{R'} = \frac{1}{f}. \tag{6.16}$$

Given R', we can compute the distance to the waist and the waist radius by Eq. (4.27), as illustrated in Fig. 6.2. If the distribution has changed from its previous condition of being either inside or outside the Rayleigh range, then the phase bias should be added or subtracted appropriately in accord with Eqs. (5.22) to (5.24). Lenses and mirrors may be defined with spherical, cylindrical, or toroidal elements. In the paraxial approximation, these elements are treated only to second order, so no spherical aberration is imposed. Phase bias is represented to second order with an R_x and R_y. If the phase bias is selected so that $f_x = R_x$ and $f_y = R_y$, there will be no phase bias explicitly in the complex array. Propagation can then be done using the effective propagation methods described in Chapter 5.

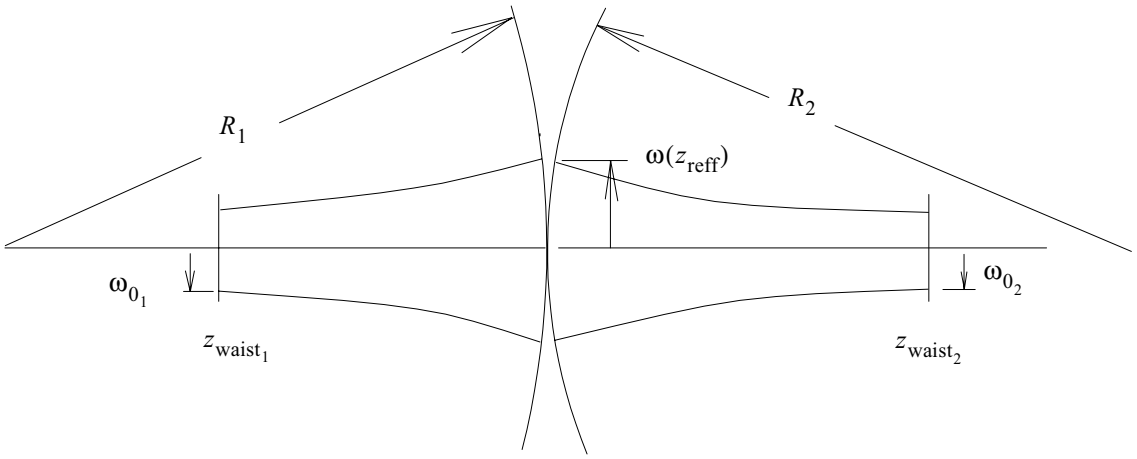


Fig. 6.2. A new gaussian waist size and location is determined by an optical element. The distribution at the element may be either inside or outside the Rayleigh distance as determined by the new gaussian waist in image space and the phase bias should be updated to reflect the new conditions.

6.1.1 Gradient Index Optics

GRIN elements may be modeled in terms of their paraxial properties in the same fashion as conventional systems. Once the geometrical paraxial behavior and the aberration properties (if any) are known GRIN elements are no more difficult to treat analytically than conventional optical systems. The paraxial ray equations take the form

$$\begin{pmatrix} y(z_2) \\ u(z_2) \end{pmatrix} = \begin{pmatrix} f_{11}(z_1,z_2)\, f_{12}(z_1,z_2) \\ f_{21}(z_1,z_2)\, f_{22}(z_1,z_2) \end{pmatrix} \begin{pmatrix} y(z_1) \\ u(z_1) \end{pmatrix}, \tag{6.17}$$

where z_1 and z_2 are the starting and ending axial positions, y and u are the paraxial ray height and angle, and $f_{ij}(z_1 - z_2)$ are functions of the index of refraction and constitute the elements of the ABCD matrix. Marchand gives expressions for tracing paraxial rays is spherical, radial, and axial gradients [2]. The solutions for lens-like media and for the Maxwell's Fisheye lens are known analytically and are listed in Table 6.1. ABCD matrices of complex systems may be built up by multiplication in the usual fashion.

6.1.1.1 Quadratic Index Variation

Quadratic index variation may be defined by the index change Δn at specified normalizing radius r_0 obeys the equation (see Siegman[1]):

$$n(r) = n_0 - \frac{1}{2}\gamma^2 r^2. \tag{6.18}$$

We have the ABCD solution:

$$\mathrm{ABCD}(z) = \begin{pmatrix} \cos(\gamma z) & \sin(\gamma z)/\gamma \\ -\sin(\gamma z)/\gamma & \cos(\gamma z) \end{pmatrix}. \tag{6.19}$$

Consider a collimated beam and a small thickness Δz of GRIN media. From Eq. (6.18) we have the wavefront error $W_1(r)$:

$$W_1(r) = -\frac{1}{2\lambda}\gamma^2 r^2. \tag{6.20}$$

For the same conditions, the ABCD operator gives the wavefront error $W_2(r)$ from the C component of Eq. (6.19). The change in angle from u = 0 for a small thickness of GRIN media is:

$$u' = -\frac{y\phi}{n_0} = yC. \tag{6.21}$$

The corresponding radius is $R'_2 = \frac{u'}{y}$ so

$$R'_2 = \frac{1}{C} \approx \frac{1}{-\gamma \sin(\gamma z)} \approx -\frac{1}{\gamma^2 z}. \tag{6.22}$$

We see that for a small thickness of GRIN media and collimated incident light the radii calculated from Eqs. (6.18) and (6.19) are equal $R'_1 = R'_1$.

Numerical examples

Consider a gradient index of the form, $n(r) = n_0(1 - Ar^2/2r_0^2)$ where $A = 0.038$, $r_0 = 0.00525$ cm, $n_0 = 1.49$, $\gamma^2 = An_0/r_0^2 = 2054.24$ cm^{-2}. For z =.0010 cm, $R = -0.4868$ cm.

$W(r_0) = r_0^2/2R = -2.831\times10^{-5}$. For $\lambda_0 = 0.000132$ cm and $n_0 = 1.49$, $W(r_0) = -0.3196\lambda$.

6.1.2 System Analysis

We shall limit our treatment to separable systems and consider only the y-z plane to simplify the discussion. Most optical systems are separable, but some, such as those having cylindrical elements at arbitrary azimuthal angles, are not. Siegman gives a thorough discussion of the use of nonorthogonal systems by a generalized ABCD matrix[1].

In paraxial optics we identify the concept of an optical space with each region of constant index. Any plane in one space has an image in each of the other spaces. In non-GRIN system, a separate optical space is defined for each region of constant index. For GRIN elements we can identify an optical space with each axial position in the medium. For either type of system, the optical space is considered to be an infinite space of constant index. Considering one optical space as the object space, we can find an image plane in any other optical space using the paraxial properties, as indicated by Fig. 6.3. These concepts are commonly applied for conventional systems. If this is counter-intuitive it may help to consider that a GRIN element can always be broken up into a sequence of simple lenses, and a conventional optical system consisting of spaces and lenses may be considered a particularly bizarre GRIN system.

6.2 Beam Propagation Method

The very simplest model of the beam propagation method (BPM) in inhomogeneous media is done using split-steps of adding the effects of inhomogeneities as aberration and collimated propagation in homogeneous media. While suitable for very weak index variations and short propagation distances, it is difficult numerically because the beam changes size in the array due to first order geometric propagation and the phase curvature may become large. Fig. 6.3a illustrates propagation through gradient media using collimated BPM. It is difficult to maintain proper sampling and large phase values are manifest in the array. In addition, it is essential to remember that the wavefront is curved and to evaluate the index of refraction of off-axis points on the curved wavefront.

A more sophisticated model using noncollimated propagation can be done by modeling the medium as a series of refractive index changes, elementary lenses, aberration, and homogeneous diffraction steps. The use of elementary lenses and noncollimated propagation greatly aids the process, as shown in Fig. 6.3b. Curved reference surfaces help to limit the phase excursions in the computer arrays and noncollimated propagation methods result maintain good sampling as the beam changes size. The ideal way to set up a series of lenses and spaces is to use the paraxial behavior as a paradigm. The paraxial ray equivalent of the noncollimated BPM operations is

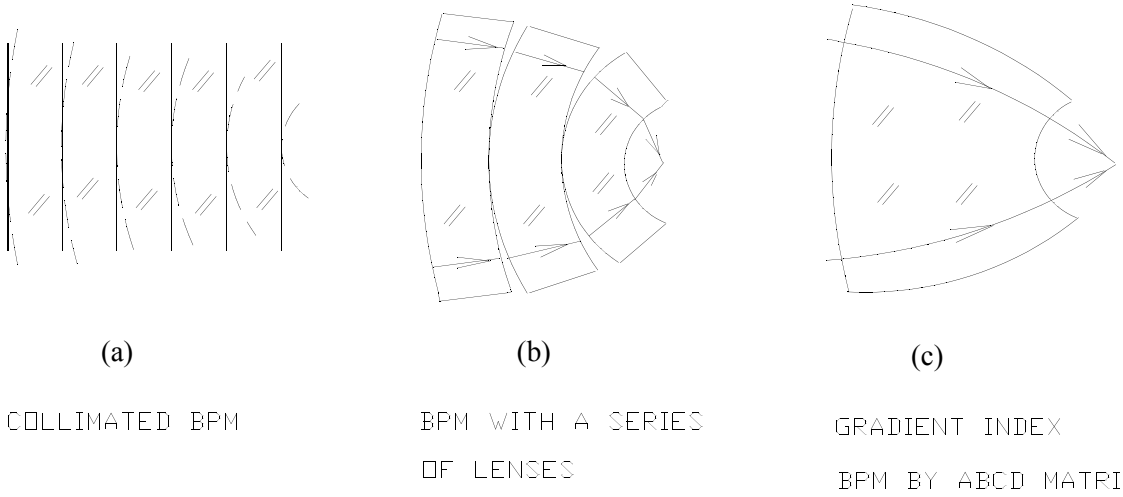


Fig. 6.3. In the simplest form, the beam propagation method (BPM)BPM method may be applied as a series of collimated propagation steps with index variations treated as aberration (a). Maintaining good sampling through the course of the propagation is difficult if the gradient effect is strong. A more sophisticated approach models gradient index effects as a series of lenses and spaces (b). This method more readily achieves good sampling but it is necessary to define the choice of lenses and spaces by some algorithm or by monitoring the sampling at each step in the propagation and determining the best optical sequence step-by-step. The ABCD method allows a step of any length to be taken with a single propagation step, irrespective of the complexity, provided the propagation may be characterized by paraxial properties (c).

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \cdots \begin{pmatrix} 1 & t_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\varphi_2/n_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & n_1/n_2 \end{pmatrix} \begin{pmatrix} 1 & t_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\varphi_1/n_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & n_0/n_1 \end{pmatrix}. \tag{6.23}$$

In more compact notation we have,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \prod_{k=1}^{N} \begin{pmatrix} 1 & t_k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\phi_k/n_k & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & n_{k-1}/n_k \end{pmatrix} \quad , \tag{6.24}$$

where the subscripts indicate the properties of the respective incremental homogeneous steps. By properly selecting the index ratio, optical power, and effective propagation length for each step, it is possible, in the limit, to make the sequence converge to the ABCD matrix. To develop the best sequence of elementary operations for implementing Eq. (6.16), we can trace a paraxial ray representing the beam size, select an axial sampling interval (which determines the variables t_1, t_2, etc.) and then calculating the powers of elements to best approximate the curve trajectory of the ray by a series of straight ray segments, as shown in Fig. 6.4.

It is obvious that a valid construction of the split step method for diffraction propagation should agree with paraxial optics. This may be accomplished by proper selection of the free variables in Eq. (6.24). One of the common tests of diffraction algorithms is to propagate a gaussian beam. If the algorithm fails to meet

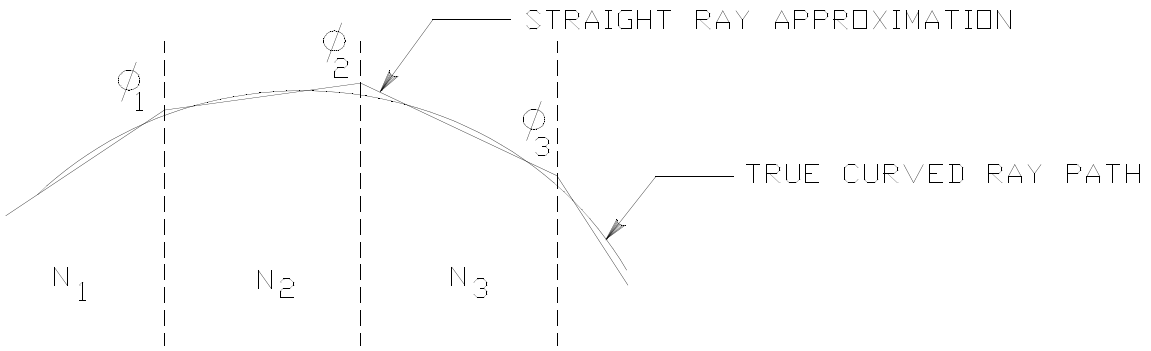


Fig. 6.4. A curved ray path may be approximated by a series of straight line segments. The power of each element is adjusted to put the ray in the next optical space to best match the curve in that segment.

the paraxial test, it will certainly not propagate a gaussian beam correctly and will therefore fail this elementary test of diffraction accuracy. The disadvantage of Eq. (6.24) is that a diffraction propagation is needed for each thickness operator, leading to a large number of propagation steps.

By replacing the paradigm of an infinite series of incremental lenses and spacings with the paraxial operator, we have the generalized BPM. The use of four free variables in Eq. (6.4) allows an exact match to paraxial behavior, even over long propagation steps. This method may be applied in exactly the same fashion for conventional and GRIN optics the only difference being in the calculation of aberrations and the details of the ABCD matrix. For stigmatic media the aberration step may be eliminated and the equivalent ABCD matrix for propagation of any length may be computed and only one propagation step used, as illustrated in Fig. 6.3c. Even for significant aberration, only a few steps may be required, as discussed more fully below.

6.3 Non-Stigmatic Effects

Extending the homogeneous diffraction equations to include all types of stigmatic media does not, of course, obviate the need to interrupt the diffraction calculation to include aberration, apertures, nonlinearities, etc. However, aberration-free GRIN media need not be considered by the split step method. For GRIN media with aberrations of the form $n(x, y, z)$, the index variations should be split into stigmatic and non-stigmatic parts of the form

$$N(x, y, z) = n_s(x, y, x) + n_e(x, y, x), \tag{6.25}$$

where n_s and n_e are the stigmatic and error components. Some judgment may be required in the selection of the stigmatic component. Given experimental data, it may be necessary to find a “best fit’ of the free variables of the selected ideal element. For example with the Maxwell's fisheye, the free variables may be the best center of symmetry (x_0, y_0, z_0), center index n_0, and focal length f with the functional form

$$N(x, y, z) = \frac{n_0}{1 + \dfrac{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}{f^2}}. \tag{6.26}$$

Presumably the residual function $n_e(x, y, x)$ is a small percentage of the stigmatic component. The aberration may be determined by integrating along the optical rays

$$W(\Delta s) = \int_0^{\Delta s} n_e(s)\mathrm{d}s\,, \tag{6.27}$$

where s indicates the ray path. This is equivalent to calculating the effect of collimated light going through a slab of glass with slight inhomogeneities. If the magnitude and spatial frequency components of the index variations are modest, the aberrations may be lumped into a single plane at the beginning of the propagation.

Even an ideal component may exhibit aberrations because of the local phase of the field which is incident on the lens. Since the local phase is partly determined by diffraction propagation, the aberration calculation should consider these local phase variations. This is true for both conventional and GRIN elements. The aberration properties of most elements are not strongly altered by slight changes of the local phase and this affect is generally neglected. In any case, the methods described in this paper neither alleviate nor exacerbate this small source of error.

6.4 High Numerical Aperture Lenses

Fresnel diffraction theory is not adequate to treat high numerical aperture objective lenses. A vector diffraction treatment will show an elliptical shape to the far-field image for a linearly polarized incident field. Vector diffraction calculations may be implemented with varying degrees of sophistication — enabling the accurate treatment of larger numerical apertures for the more sophisticated models. See Fig. 6.5. The method employed by the command HIGNA is based on dipole projection. The incident field is assumed to have minimum angular spread and to be well represented by the two transverse components of the electric field,

$$\text{incident field} = E_x\hat{\boldsymbol{x}} + E_x\hat{\boldsymbol{y}}\,. \tag{6.28}$$

The effect of refraction is to redirect the optical ray so that it heads toward the image point. The electric field components now have the form,

$$\text{refracted field} = E_x\hat{x}' + E_x\hat{y}' + 0\hat{z}'\ , \tag{6.29}$$

where $\boldsymbol{x}'$, $\boldsymbol{y}'$, and $\boldsymbol{z}'$ represent the coordinate system along the refracted ray. The optical fields at the image are taken with respect to the unrotated coordinate system designated by $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$. We may write the refracted field as,

$$\text{refracted field} = [(\hat{\boldsymbol{x}}')(\hat{\boldsymbol{y}}')(\hat{\boldsymbol{z}}')]\begin{pmatrix} E_x \\ E_y \\ 0 \end{pmatrix} = \mathbf{R}[(\hat{\boldsymbol{x}})(\hat{\boldsymbol{y}})(\hat{\boldsymbol{z}})]\begin{pmatrix} E_x \\ E_y \\ 0 \end{pmatrix}. \tag{6.30}$$

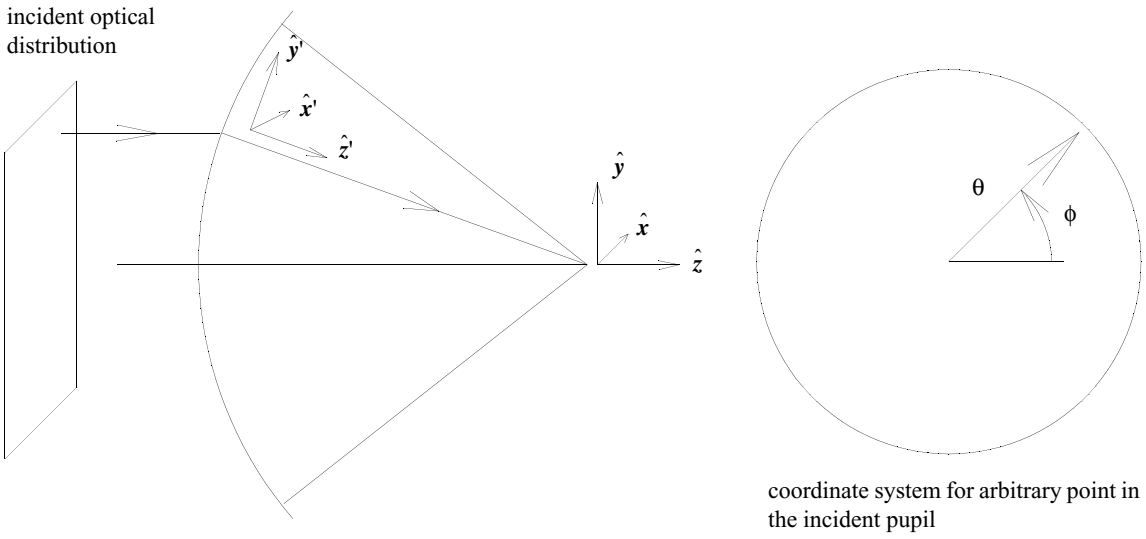


Fig. 6.5. Sketch of dipole projection theory of vector diffraction imaging. A distribution is incident from the left to a high numerical aperture lens. Consider dipoles in the $\hat{x}$, $\hat{y}$, and $\hat{z}$ directions for the incident radiation. After refraction, the coordinate system is rotated, taking the form, $\hat{x}'$, $\hat{y}'$, and $\hat{z}'$. The primed system projects onto the original system causing mixing of polarization states, including projection of the transverse components onto the $\hat{z}$ axis. Each point in the pupil is designated by the radial and azimuthal angles, θ and ϕ respectively.

where **R** is the rotation matrix. The projection of E_x and E_y onto the unrotated coordinate system determines the effect on the image distribution. The electric components of the refracted radiation in the unrotated coordinate system are

$$\text{refracted field} = \boldsymbol{R}\begin{pmatrix} E_x \\ E_y \\ 0 \end{pmatrix} = \begin{pmatrix} E_x' \\ E_y' \\ E_z' \end{pmatrix}. \tag{6.31}$$

The matrix $\boldsymbol{R}$ is the rotational matrix which transforms the initial coordinate system to the rotated coordinate system. This matrix may be found by expressing the transverse position of the ray in the pupil in terms of the polar angles θ and φ such that

$$\theta = \sin^{-1}\left(\frac{r}{f}\right), \varphi = \tan^{-1}\left(\frac{y}{x}\right), \tag{6.32}$$

where $r = \sqrt{x^2 + y^2}$ and f is the focal length. The rotational matrix is found from

$$\boldsymbol{R} = \boldsymbol{R}_z(\varphi)\boldsymbol{R}_y(\theta)\boldsymbol{R}_z^{-1}(\varphi). \tag{6.33}$$

Rotations about the x-, y-, and z-axes are defined to be

$$\boldsymbol{R}_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix}, \tag{6.34a}$$

$$\boldsymbol{R}_y(\beta) = \begin{pmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{pmatrix}, \tag{6.34b}$$

$$\boldsymbol{R}_z(\gamma) = \begin{pmatrix} \cos(\gamma) & -\sin(\gamma) & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{6.34c}$$

The composite rotation is

$$\boldsymbol{R}_z(\varphi)\boldsymbol{R}_y(\theta)\boldsymbol{R}_z{}^{-1}(\varphi) = \begin{pmatrix} \cos(\theta)\cos^2(\varphi) + \sin^2(\theta) & (\cos(\theta)-1)\cos(\varphi)\sin(\varphi) & \sin(\theta)\cos(\varphi) \\ (\cos(\theta)-1)\cos(\varphi)\sin(\varphi) & \cos(\theta)\sin^2(\varphi) + \cos^2(\varphi) & \sin(\theta)\sin(\varphi) \\ -\sin(\theta)\cos(\varphi) & -\sin(\theta)\sin(\varphi) & \cos(\theta) \end{pmatrix}. \tag{6.35}$$

6.5 Optical Systems

Geometrical optics design is used to correct the aberrations of a system for a specific object-image conjugate pair. Consider Fig. 6.6, geometrical analysis is capable of calculating the aberrations of the system. Presumably, the system is relatively well-corrected at the image point, but the aberration may be large at intermediate image points. Large but compensating aberrations are easily handled in geometrical optics but not in physical optics, where the wavefront slopes must be resolved at all points. The best method of modeling such systems, is to combine geometrical and physical optics methods. The system is ray traced to find the aberrations and the paraxial behavior, including the location and size of the entrance pupil, as shown in the lower schematic of Fig. 6.6. The equivalent system may be represented by operators where only the B elements indicate diffraction steps. Assuming the calculation begins at the object and the entrance pupil is a distance t in front of the object, the system may be represented by

$$\text{system ABCD operator} = \begin{pmatrix} M & 0 \\ -\varphi & 1/M \end{pmatrix}\begin{pmatrix} 1 & -t \\ 0 & 1 \end{pmatrix}(\text{ABERRATION})(\text{CLAP})\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \tag{6.36}$$

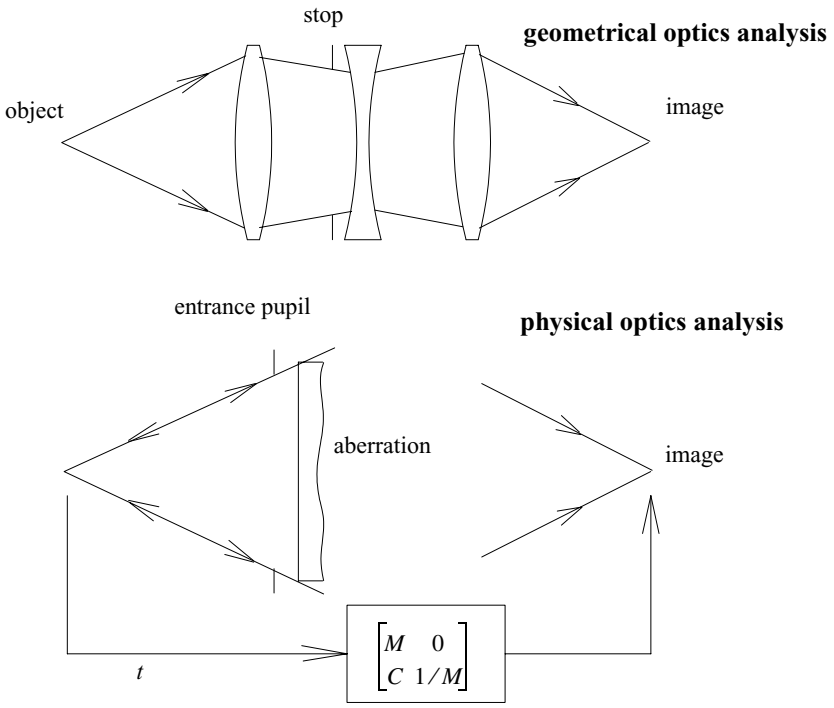


Fig. 6.6. Consider a schematic of an optical system as shown above. In geometrical analysis, we trace rays from the object to the image and calculate the aberrations. Each component may introduce substantial aberration, but in the well-corrected system all aberrations cancel for the complete system. For physical optics, the large amount of aberration due to the individual elements induces aliasing in the intermediate optical spaces. In physical optics it is more accurate and much faster to represent the system by an optical equivalent consisting of an entrance pupil, aberration, and an ABCD operator representing the transformation from object plane to image plane. We propagate from the object to the entrance pupil, apply an aperture representing the entrance pupil size, aberration, propagate back to the object, and then transform from object to image planes.

where φ is the optical power of the system. By inspecting the B-terms, we see that the only propagation steps are t and $-t$. If the clear aperture and aberration are omitted, then these steps exactly cancel so we will see excellent numerical behavior. Even with an aperture of finite size and aberration typical of a corrected optical system, the numerical performance will be excellent.

Information to be transferred:

- ABCD matrix, abcd/operator,
- optical path length (OPL),
- coordinate transformation, global/operator,
- aberration added at exit pupil, abr/zern/number,
- clipping apertures, superimposed or separated.

References

1. A. Siegman, *Lasers*, Chap. 20, p. 805, University Science Books, Mill Valley, CA, (1986).

2. E. W. Marchand, *Gradient Index Optics*, Academic Press, New York (1978).

3. S-H Hwang and G. Lawrence, “Beam propagation in gradient refractive-index media,” Appl. Opt., Vol. 31, p5201-5210 (1992).

4. A. Yariv, *Introduction to Optical Electronics*, 2nd Edition, Chap. 2, Holt, Rinehart and Winston (1971).

7. Aberration

In the previous sections we have treated the optical system as ideal without consideration of aberration or apertures. Proper treatment of aberrations is one of the most important application of optical modeling methods. Aberration may be introduced by the intrinsic design of the optical configuration, by manufacturing errors, by inhomogeneous index of refraction, self-focusing, thermal blooming, and other effects. Aberration may be removed by spatial filters and other aperturing effects.

7.1 Aberration Polynomials

The phase errors of lenses, mirrors, windows, and media of variable index of refraction can be represented by aberration polynomials. The aberration is introduced into the complex amplitude distribution by

$$a(x, y) \rightarrow a(x, y)e^{jkW(x, y)}, \tag{7.1}$$

where k is the wave number and $W(x, y)$ is the aberration function.

The classical Seidel polynomials are much used in the study of rotationally symmetrical optical systems and are useful for low order aberration. If they are generalized to allow arbitrary azimuthal orientation, they also do a good job of modeling simple component aberration due to misalignment or distortion of an element. These aberrations have the form listed in Table 7.1. Figure 7.1a shows spherical aberration. Figure 7.1b shows a combination of spherical aberration and astigmatism.

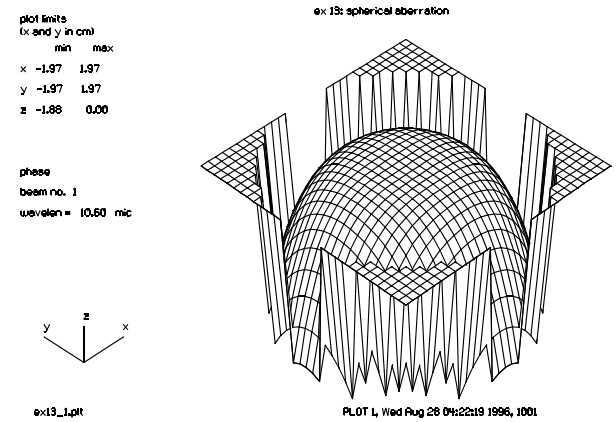


Fig. 7.1.a. Spherical aberration of a circular pupil function shown in the form of an isometric plot.

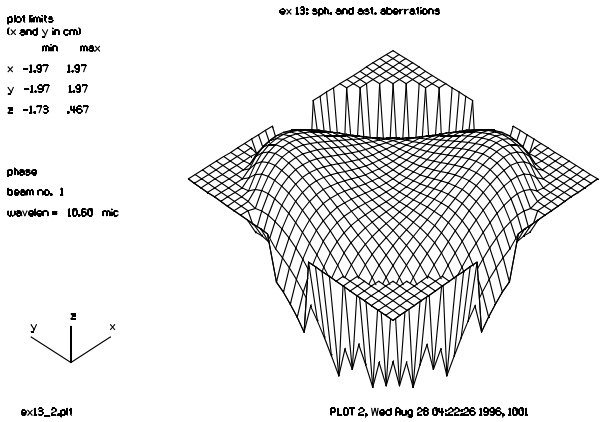


Fig. 7.1b. Spherical aberration (the same as Fig. 7.1a) plus astigmatism on a circular aperture.

Also, listed in Table 7.1 is the general form of the Zernike aberrations which are frequently used to characterize manufacturing error or wavefront error. Zernike aberrations may be calculated from interferometric data using readily available computer software. Fig. 7.2a shows Zernike aberration of a mixture of high order polynomials with the same aberration shown as a contour plot in Fig. 7.2b.

Table. 7.1. Seidel and Zernike aberration polynomials. r and θ are the pupil coordinates, r_n is the radius for normalization of the pupil.

Seidel aberration	
Tilt	$W(r,\theta) = W_{11}\left(\frac{r}{r_{\text{norm}}}\right)\cos(\theta-\theta_0)$
Focus	$W(r,\theta) = W_{20}\left(\frac{r}{r_{\text{norm}}}\right)^2$
Astigmatism	$W(r,\theta) = W_{22}\left(\frac{r}{r_{\text{norm}}}\right)^2\cos^2(\theta-\theta_0)$
Coma	$W(r,\theta) = W_{31}\left(\frac{r}{r_{\text{norm}}}\right)^3\cos(\theta-\theta_0)$
Spherical	$W(r,\theta) = W_{40}\left(\frac{r}{r_{\text{norm}}}\right)^4$
Zernike aberration	
Radial Zernike	$W(r,\theta) = R_n(r)$
Sine Zernike	$W(r,\theta) = R_{nm}(r)\sin(m\theta)$
Cosine Zernike	$W(r,\theta) = R_{nm}(r)\cos(m\theta)$

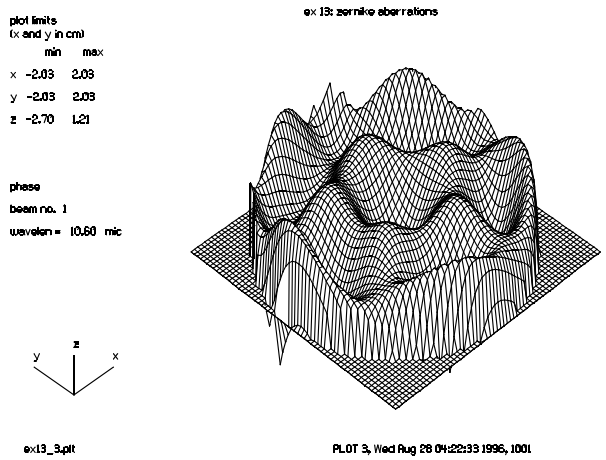


Fig. 7.2.a. Higher order Zernike polynomial aberration.

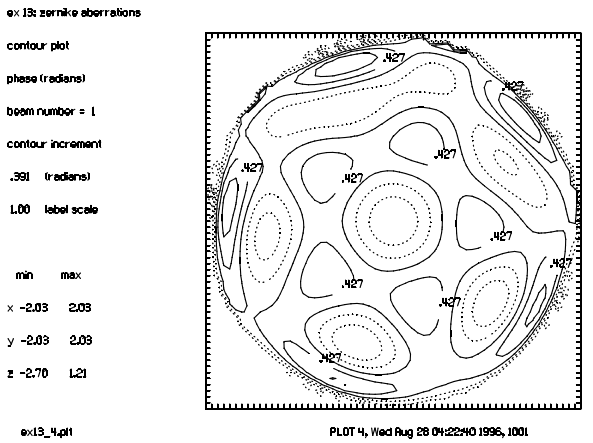


Fig. 7.2b. Higher order Zernike polynomial aberration.

7.2 Random Aberration

Very frequently the exact characteristics of the aberration are not known. We may not have information on the manufacturing error of components either because they have not been measured or the design process is still in the process of definition and only the specifications are known. In either case, the modeling may

proceed on the basis of smoothed random wavefronts which are constructed to meet the specifications. Atmospheric turbulence is a special type of random wavefront, which conforms to certain special statistics.

7.2.1 Generation of Random Aberration

The random wavefront should have the requisite autocorrelation radius and RMS or peak-to-valley wavefront error. For a stationary process, i.e., one whose statistics are independent of spatial coordinates, the autocorrelation function is

$$R(\Delta x, \Delta y) = \langle F(x, y)F(x + \Delta x, y + \Delta y)\rangle . \tag{7.2}$$

The characteristic width of the autocorrelation function is called the autocorrelation width. If the phase is relatively slowly varying across the aperture the autocorrelation width is large. To construct the random wavefront with the correct autocorrelation function, we first construct a delta-correlated wavefront, i.e., one with no correlation between points in the array,

$$R_{\Delta}(\Delta x, \Delta y) = \langle W_{\delta}(x + \Delta x, y + \Delta y)W_{\delta}(x, y)\rangle = \sigma^2\delta(\Delta x, \Delta y), \tag{7.3}$$

where σ^2 is the wavefront variance and $W_{\delta}(x, y)$ is a delta-correlated array

This can be done by generating random numbers for each point in the array in a manner illustrated in Fig. 7.3. We can smooth $W_{\delta}(x, y)$ to meet the required specifications. The smoothing operation is

$$W(x, y) = W_{\delta}(x, y)**s(x, y), \tag{7.4}$$

where $s(x, y)$ is a normalized smoothing function and $**$ represents two-dimensional convolution. The autocorrelation function of the smoothed wavefront, $W(x, y)$ is

$$R(\Delta x, \Delta y) = \left\langle \iint W(x, y)W(x + \Delta x, y + \Delta y)\mathrm{d}x\mathrm{d}y\right\rangle , \tag{7.5}$$

$$R(\Delta x, \Delta y) = \left\langle \iiiiii W_{\delta}(x_1, y_1)s(x - x_1, y - y_1)W_{\delta}(x_2, y_2)s(x + \Delta x - x_2, y + \Delta y - y_2)\mathrm{d}x_1\mathrm{d}y_1\mathrm{d}x_2\mathrm{d}y_2\mathrm{d}x\mathrm{d}y\right\rangle . \tag{7.6}$$

Moving the statistical averaging inside the integral to address only the stochastic terms we have

$$R(\Delta x, \Delta y) = \iiiiii \langle W_{\delta}(x_1, y_1)W_{\delta}(x_2, y_2)\rangle s(x - x_1, y - y_1)s(x + \Delta x - x_2, y + \Delta y - y_2)\mathrm{d}x_1\mathrm{d}y_1\mathrm{d}x_2\mathrm{d}y_2\mathrm{d}x\mathrm{d}y , \tag{7.7}$$

$$R(\Delta x, \Delta y) = \iiiiii \sigma^2\delta(x_1 - x_2, y_1 - y_2)s(x - x_1, y - y_1)s(x + \Delta x - x_2, y + \Delta y - y_2)\mathrm{d}x_1\mathrm{d}y_1\mathrm{d}x_2\mathrm{d}y_2\mathrm{d}x\mathrm{d}y , \tag{7.8}$$

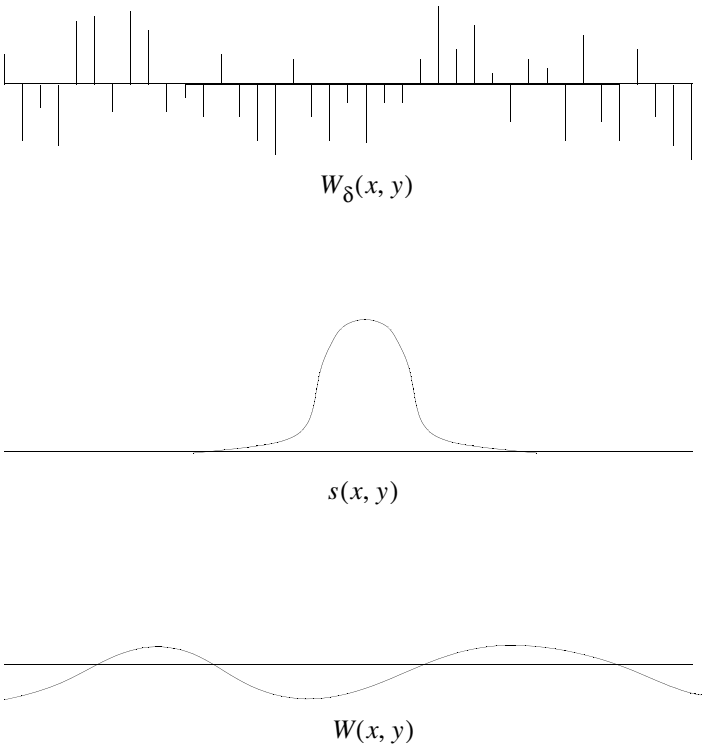


Fig. 7.3. The random wavefront is constructed from a delta-correlated wavefront generated from a sequence of random numbers. The delta-correlated wavefront is smoothed to create a wavefront with the desired statistics.

$$R(\Delta x, \Delta y) = \iint \sigma^2 s(x, y) s(x + \Delta x, y + \Delta y) \mathrm{d}x \mathrm{d}y = \sigma^2 s(x, y) \star\star s(x, y), \tag{7.9}$$

where $\star\star$ represents two-dimensional complex autocorrelation. Selecting the smoothing function to have an autocorrelation function radius of a_0 (and to be normalized) to be

$$R(\Delta x, \Delta y) = \sigma^2 e^{-r^2/a_0^2} \rightarrow s(x, y) = \sqrt{\frac{4}{\pi a_0^2}} e^{-2r^2/a_0^2}. \tag{7.10}$$

Figures 7.4.a and 7.4b show random aberration created forming a set of completely uncorrelated phase points and then smoothing to the desired autocorrelation function. Figure 7.4.a has a relatively broad autocorrelation radius and Fig. 7.4b shows an autocorrelation diameter of half the size of Fig. 7.4.a. Random aberration can be used to simulate the effects of manufacturing aberration. The amplitude of the aberration and the autocorrelation width can be adjusted to fit the type of aberration expected from the manufacturing process if it is known or to fit specification values.

7.2.2 Atmospheric Aberration

The atmospheric turbulence model is based on an angular spectrum model which is readily implemented and which is readily interfaced with the adaptive optics model. Hufnagel and Stanley established a Fouier

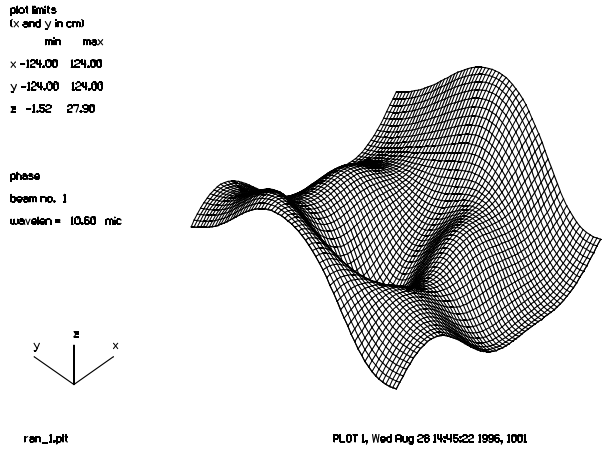


Fig. 7.4.a. The random wavefront is constructed from a delta-correlated wavefront generated from a sequence random numbers. The delta-correlated wavefront is smoothed to create a wavefront with the desired statistics.

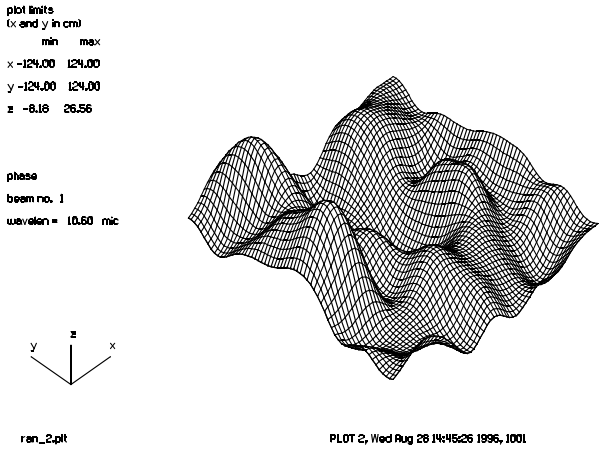


Fig. 7.4b. Random aberration formed with a narrow autocorrelation width. The aberration was formed from the same set of random numbers as Fig. 7.4.a, but was given less smoothing. By controlling the amplitude of the aberration and the degree of smoothing, the type of aberration may be controlled.

optics method of analysis[1]. Hogge and Butts approached the same subject through Zernike polynomials[2]. Fried made significant contributions including the seeing constant commonly known as Fried's parameter[3]. Rodier has provided a general overview of the subject[4]. Issue of inner and outer scale have been addressed by Lutomirski and Yura[5]. More recent work by Krapels, et. al., specifically addresses modeling[6] and by Sadot describes experimental valdiation[7]. There has been a lack of agreement among atmospheric experts with regard to proper treatment of atmospheric outer scale. This leads to uncertainty in atmospheric tilt and piston error which is problematic for uncorrected systems and for large phased arrays but not for adaptively corrected systems.

Atmospheric aberration is described by the Kolmogorov spectral distribution. The aberration is described by an integral over the altitude in the atmosphere as described by Roddier[4]

$$W^2(\rho) = \frac{0.38}{\lambda^2 \rho^{11/3}} \int_{h_{\min}}^{h_{\max}} C_n^2(h)\mathrm{d}h, \tag{7.11}$$

where $C_n^2(h)$ is the refractive index structure constant of the atmosphere as a function of altitude, h is the altitude in meters, ρ is the spatial frequency variable, λ is the wavelength, and $W^2(\rho)$ is the power spectrum of the wavefront. We may characterize the aberration by the autocorrelation diameter, r_0, sometimes known as Fried's parameter

$$r_0 = \left(0.423 k^2 \int_{h_{\min}}^{h_{\max}} C_n^2(h)\mathrm{d}h \right)^{-3/5}. \tag{7.12}$$

Fried's parameter may then be used to calculate the wavefront

$$W^2(\rho) = \frac{0.023}{r_0^{5/3}\rho^{11/3}}. \tag{7.13}$$

The inner and outer scale may be included according to Lutomirski and Yura[5],

$$W^2(\rho) = \frac{0.023e^{-\rho L_i^2}}{\left(\rho^2 + \frac{1}{L_o^2}\right)^{11/3}}, \tag{7.14}$$

where L_o is the outer scale, and L_i is the inner scale These parameters are in radians, meters, and inverse meters respectively. L_o is generally considered to be several meters and L_i to be on the order of centimeters or less.

From the relationships described above, we can see that it is possible to characterize the atmospheric aberration by Fried's parameter. It is natural therefore to separate the problem of atmospheric characterization into three parts: 1) determination of the effects of various levels of atmosphere in terms of r_0 values for each level, 2) summation of the r_0 values, and 3) implementation of aberration using the r_0 parameter. The summation of the r_0 values over different levels takes the form

$$r_{\text{total}} = (r_{0_1}^{-5/3} + r_{0_2}^{-5/3} + \ldots)^{-3/5}. \tag{7.15}$$

Propagation over large distances is accomplished by taking short steps of simple diffraction alternating with short steps of aberration introduction. Figure 7.5 shows a typical atmospheric aberration phase sheet.

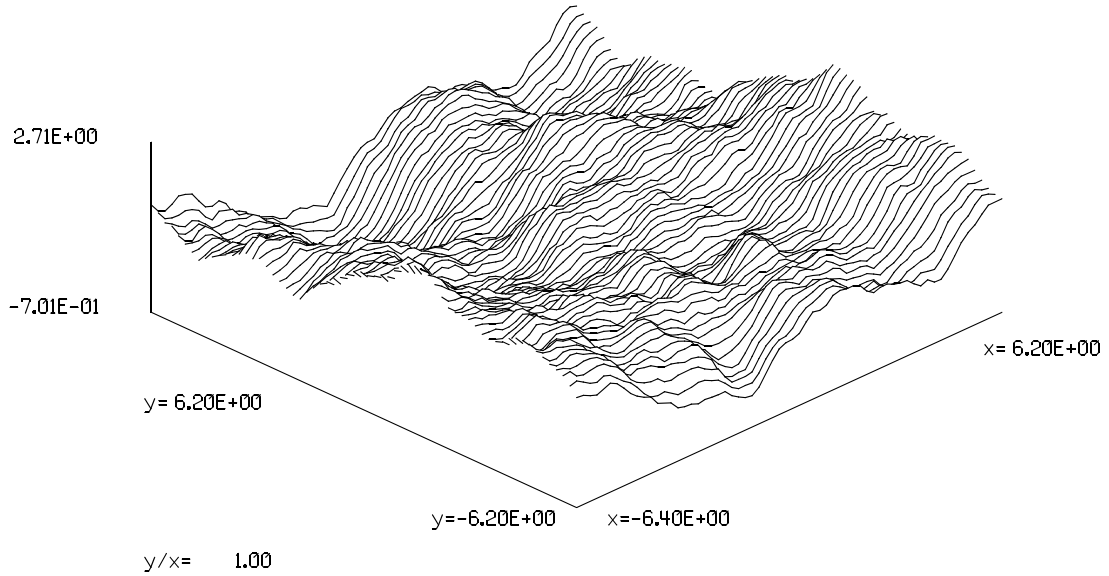


Fig. 7.5. Schematic of propagation through the air where aberration and diffraction propagation must be considered. The propagation is done in relatively short steps with the diffraction propagation and addition of aberration alternated to achieve an approximation to a continuous process.

7.3 Beam Quality Measurement

Beam quality can be measured in a variety ways. There is no method which is best for all applications. The user should carefully consider the requirements of the design and then select the method which is most appropriate; or devise a new method specifically suited to the task at hand.

7.3.1 Strehl Ratio

The Strehl ratio is commonly defined as the ratio of the peak image intensity in the far-field with aberrations to the peak image intensity in the far-field without aberration. In GLAD Strehl ratio may be calculated with the `strehl` command. It is convenient to use a modified version of Strehl ratio consisting of the ratio of the center of the far-field intensity with aberration to the far-field intensity with no aberrations. The far-field intensity is

$$I(x,y)_{\text{aberr}} = \frac{1}{\lambda^2 f^2}\left|\iint a(x,y)e^{j2\pi(x\xi+y\eta)}\mathrm{d}x\mathrm{d}y\right|^2. \tag{7.16}$$

The far-field intensity of the same intensity distribution, without aberrations is

$$I(x,y)_{\text{noaberr}} = \frac{1}{\lambda^2 f^2}\left(\iint |a(x,y)|e^{j2\pi(x\xi+y\eta)}\mathrm{d}x\mathrm{d}y\right)^2. \tag{7.17}$$

Evaluating these at $\xi = 0$ and $\eta = 0$ the Fresnel kernel disappears, and we have

$$\text{Strehl ratio} = \frac{I(0,0)_{\text{aberr}}}{I(0,0)_{\text{noaberr}}} = \frac{\left|\iint A(x,y)\mathrm{d}x\mathrm{d}y\right|^2}{\left(\iint |A(x,y)|\mathrm{d}x\mathrm{d}y\right)^2}. \tag{7.18}$$

This expression is readily evaluated by integration across the complex amplitude of the beam, and does not explicitly require propagation to the far-field.

7.3.2 Wavefront Variance

Another common measure of performance is the wavefront variance, calculated in GLAD with the `variance` command

$$\sigma^2 = \frac{\iint W(x,y)^2\mathrm{d}x\mathrm{d}y}{\iint \mathrm{d}x\mathrm{d}y} - \left(\frac{\iint W(x,y)\mathrm{d}x\mathrm{d}y}{\iint \mathrm{d}x\mathrm{d}y}\right)^2. \tag{7.19}$$

The above expression is appropriate for an aperture which is either 100% or 0% transmitting but does not treat the general complex amplitude distribution.

A well-known relationship between Strehl ratio and wavefront variance is

$$\mathrm{SR} \approx e^{-4\pi^2\sigma^2} \text{ and } \sigma^2 \approx -\frac{\ln(\mathrm{SR})}{4\pi^2}. \tag{7.20}$$

7.3.3 Optical Transfer Function

The Optical Transfer Function (OTF) is the Fourier transform of the image irradiance or, equivalently the autocorrelation of the pupil function. In GLAD, use the OTF command, as illustrated in Ex. 65 and shown in Figs. 7.6.a and 7.6b. The units of the OTF are cycles/cm.

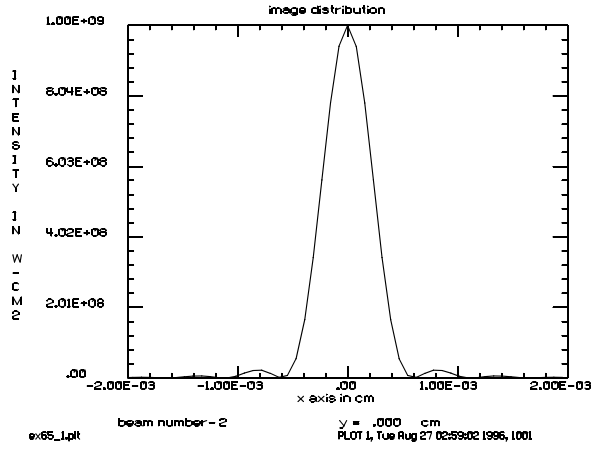


Fig. 7.6.a. Diffraction-limited image of f/5 beam.

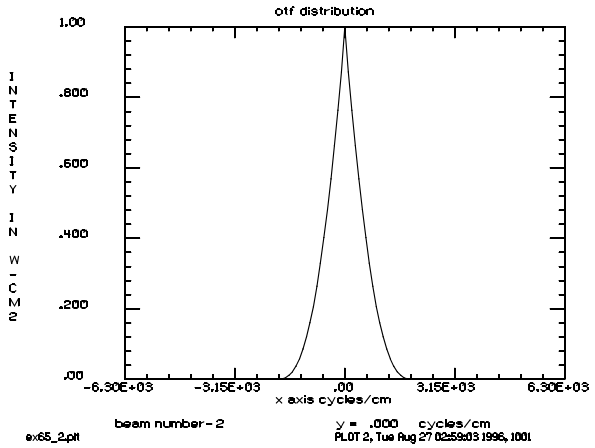


Fig. 7.6b. OTF of f/5 image.

$$\mathrm{OTF}(\xi, \eta) \propto \mathbf{FF}[i(x', y')], \tag{7.21}$$

where $i(x', y')$ is the image irradiance.

Alternatively, we may evaluate the OTF in the pupil using the normalized autocorrelation:

$$\mathrm{OTF}(\xi, \eta) = \frac{f(x, y) ** f(x, y)}{\iint |f(x, y)|^2 \mathrm{d}x\mathrm{d}y}, \quad \xi = \frac{\lambda f}{x}, \eta = \frac{\lambda f}{y}. \tag{7.22}$$

The OTF is very commonly used in image evaluation in lens design.

7.3.4 Autocorrelation Function

The autocorrelation function is useful to determine the statistical characteristics of a random speckle pattern. Fig. 7.7 and 7.8 illustrate the typical form of the autocorrelation function. Both the typical speckle size and the typical DC level may be determined. The `autocorrelation` command implements this function, as illustrated Ex. in 88f.

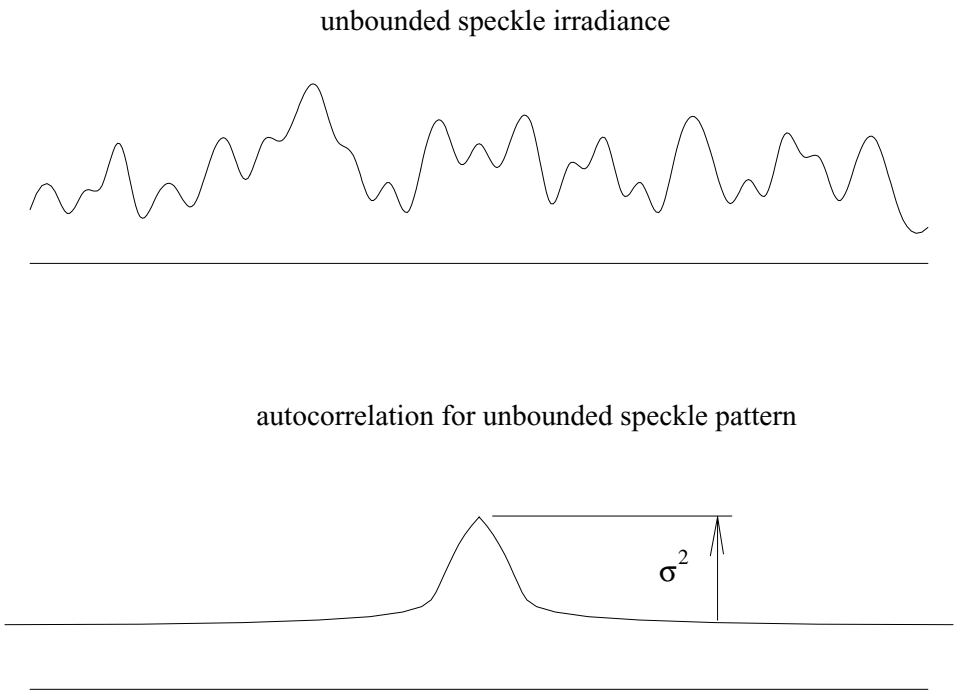


Fig. 7.7. The autocorrelation function of an unbounded speckle pattern has a DC level determined by the σ^2, the standard deviation of irradiance nonuniformity and a bump in the center which is determined by the typical speckle size.

7.3.5 Irradiance Uniformity

The degree of irradiance nonuniformity is of importance in photolithographic illumination systems using coherent or partially coherent light. Nonuniformity may be calculated with the `uniformity` command, according to

$$\text{nonuniformity (percent)} = \frac{1}{\langle I \rangle}\left(\langle I^2 \rangle - \langle I \rangle^2\right)^{1/2} \times 100\%, \tag{7.23}$$

where $\langle\ \rangle$ indicates a spatial average and the irradiance is I. This definition of irradiance nonuniformity can give a value greater than 100%, A random speckle pattern has a nonuniformity of about 100%. Fig. 7.9 illustrates a random pattern that is 100% modulated. Fig. 7.10 shows the result of summation of 100 statistically identical patterns.

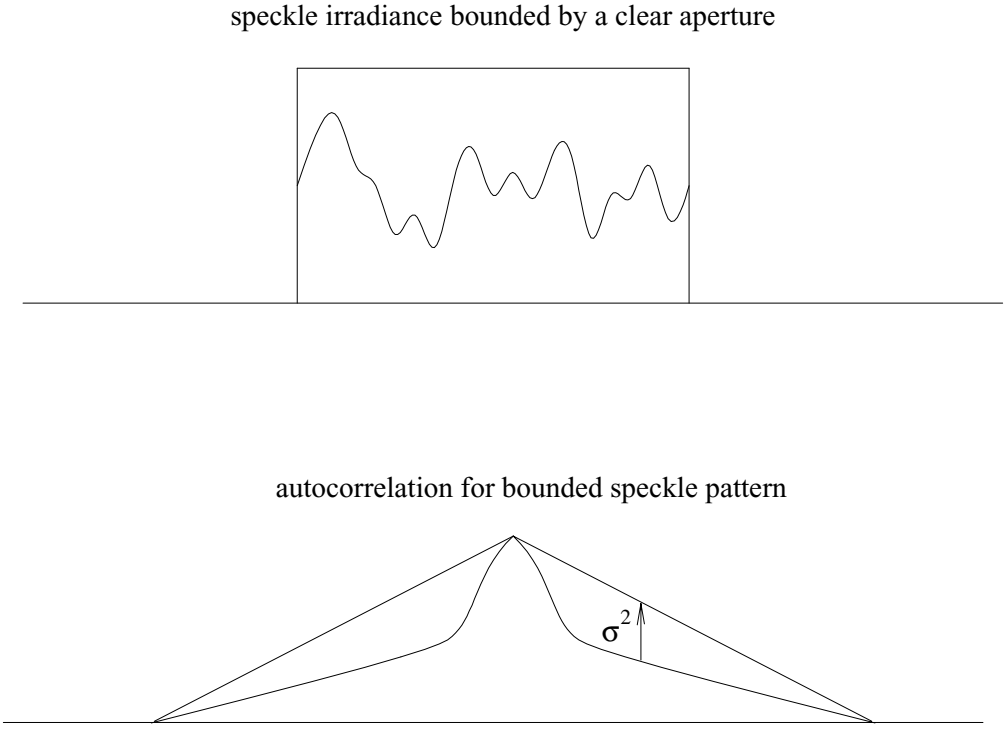


Fig. 7.8. For a finite size clear aperture the irradiance nonuniformity manifests itself as a drop from the autocorrelation of the uniformly filled aperture.

7.4 Beam Size Measurement

Beam size in the far-field is often used as a measure of beam quality. This may seem a logical choice for measurement, if the focusability of the beam is important to the application of interest. Some care must be exercised since beam size is both difficult to measure (either experimentally or numerically). A second reason is that for near-diffraction limited systems, the size of the central lobe of the diffraction pattern is relatively insensitive to the aberrations in the pupil. Phase aberrations in the pupil cause scattering into the outer rings of the far-field diffraction pattern. The central lobe decreases in height, according to the energy loss into the outer rings, but does not appreciably change width. It is for this reason, more appropriate to measure the energy within a region corresponding to the size of the central lobe rather than the width of the central lobe. Both Strehl ratio and power-in-the-bucket (PIB) are measurements of this type. It is regrettable that some experimentalists still report in the journals the width of the central lobe of their system to describe their beam quality. They inevitably find that the full-width-half-maximum (FWHM) of their beam is the same as the diffraction-limited spot and are pleased and surprised to conclude that their beam is, therefore, diffraction-limited.

As with any single number of performance, beam width does not provide a complete description of the beam. Single number measurements may be used to determine pass-fail conditions, but are not useful to diagnose the nature of the aberrations which may be present. Measurement of beam width is generally not

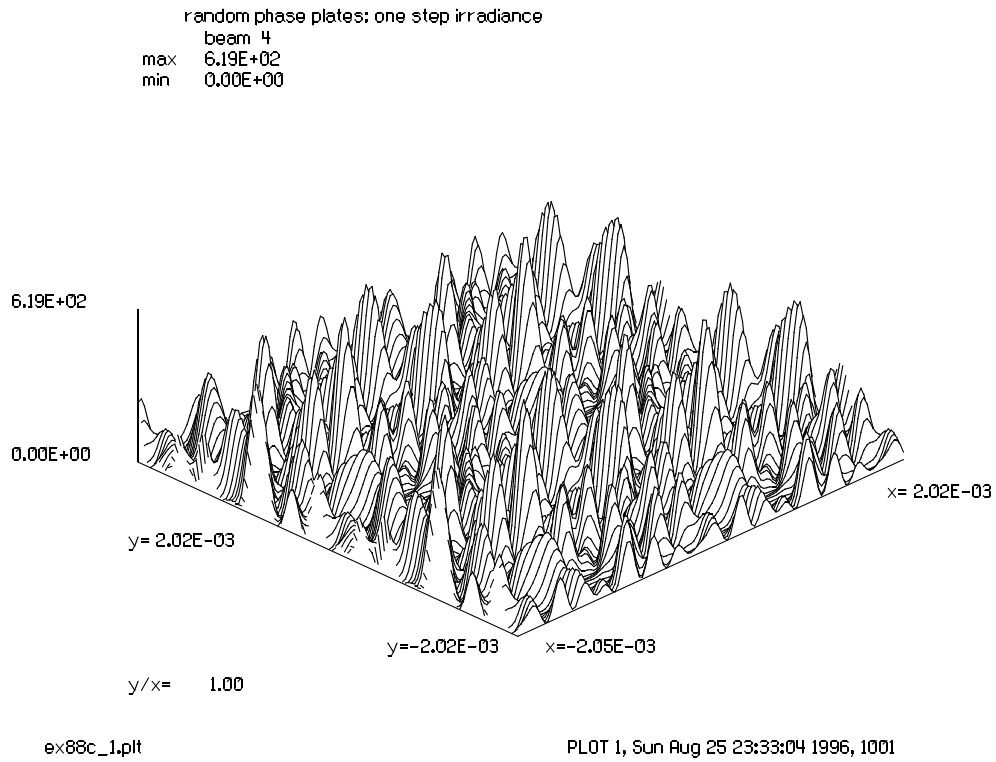


Fig. 7.9. A speckle pattern with 100% nonuniformity. This distribution was calculated by Ex88c.inp.

helpful in determining what corrections should be made to the system. They also are not very useful for alignment. The most powerful method of diagnosis is interferometry to determine the phase aberrations in the pupil. It may seem more direct to make measurements in the far-field, if we wish to specify far-field performance. However, defects in the far-field irradiance are very often due to phase aberrations in the pupil. It is quite difficult to determine phase aberrations from image irradiance, because the vital phase information is lost by the square law detector used to measure irradiance. Hence, interferometry is greatly preferred for diagnosis. Ideally, the irradiance in the pupil plane should also be measured to completely determine the complex amplitude state of the beam. From this information, and using the diffraction capabilities in GLAD all aspects of the beam may be determined.

If one is determined to measure the beam size, then an appropriate definition of beam size must be established. Seigman has recommended a measure called M^2. This is essentially the same as measuring the second spatial moment irradiance:

$$\sigma_x^2 = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(x,y)(x-\bar{x})^2 \mathrm{d}x\mathrm{d}y}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(x,y)\mathrm{d}x\mathrm{d}y}, \tag{7.24}$$

where $\bar{x}$ is the centroid. The standard deviation in the spatial domain σ_x and the comparable standard deviation in the frequency domain:

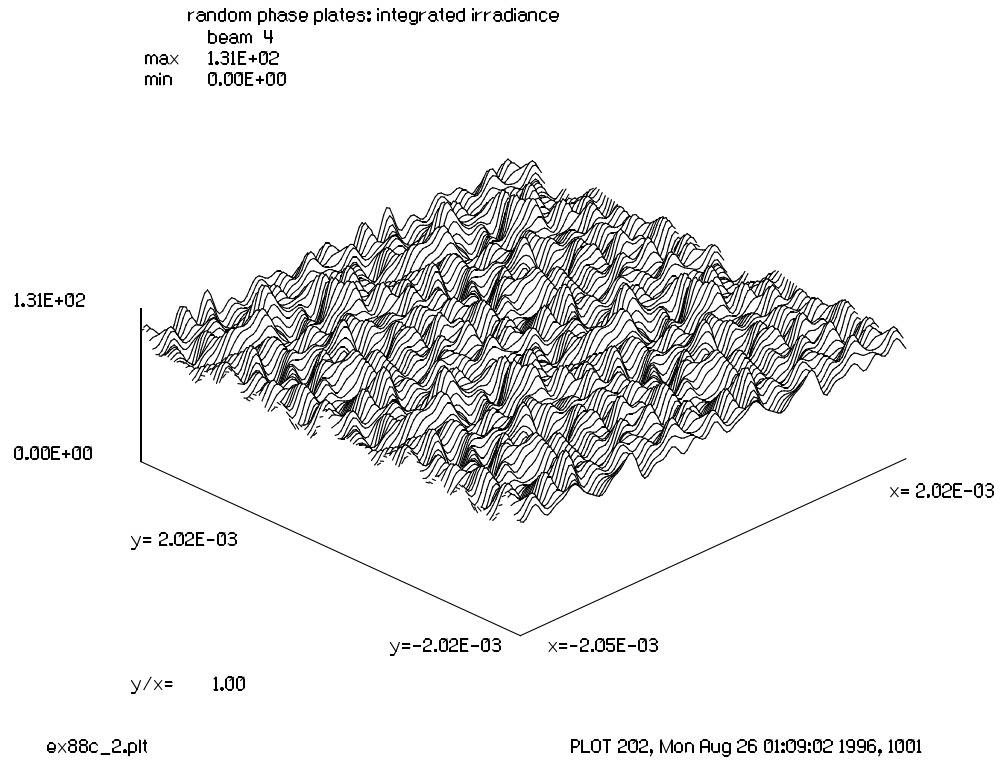


Fig. 7.10. A speckle pattern with 10% nonuniformity due to the incoherent summation of 100 different speckle patterns, from Ex88c.inp.

$$\sigma_{s_x}^2 = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(s_x, s_y|)(s_x - \overline{s_x})^2 \mathrm{d}s_x \mathrm{d}s_y}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(s_x, s_y|)\mathrm{d}s_x \mathrm{d}s_y}. \tag{7.25}$$

We find M^2 from

$$M^2 = 4\pi\sigma_x\sigma_{x_x}. \tag{7.26}$$

For a perfect gaussian beam we have $M^2 = 1$ and $\sigma_x = \omega/2$. For imperfect beams $M^2 > 1$. This definition is based on calculation of standard deviation — an easy task for well-behaved analytical functions, most particularly the Hermite-gaussians — but a very difficult task for real beams because of the susceptibility to noise. Noise is inevitable in both numerical calculations and physical measurements. Consider the addition of a uniform noise source to the calculation of variance:

$$\sigma_{x_T} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(i(x,y)+i_N(x,y))(x-\bar{x})^2 \mathrm{d}x\mathrm{d}y}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(i(x,y)+i_N(x,y))\mathrm{d}x\mathrm{d}y} = \sigma_x^2 + \sigma_N^2. \tag{7.27}$$

Unfortunately, $\sigma_N^2 \to \infty$ for any finite value of noise, because of the weighting by x^2. The quadratic weighting is such a strong effect that it makes calculation of the variance extremely inaccurate.

References

1. R. E. Hufnagel and N. R. Stanley, "Modulation transfer function associated with image transmission trough turbulent media," J. Opt. Soc. Am., **54**, 52–61 (1964).
2. C. Barry Hogge and R. Russell Butts, "Frequency Spectra for Geometric Representation of Wavefront Distortions Due to Atmospheric Turbulence," IEEE Transactions on Antennas and Propagation, Vol. AP-24, No. 2, 144–154 (March, 1976).
3. D. L. Fried, "Optical resolution through a randomly inhomogenous medium for very long and very short exposure times," J. Opt. Soc., Am, **56**, 1372–1379 (1966).
4. F. Roddier, "The Effects of Atmospheric Turbulence in Optical Astronomy," E. Wolfe, Ed., Progress in Optics XIX, North-Holland (1981).
5. R. F. Lutomirski and H. T. Yura, "Aperture Averaging Factor for a Fluctuating Light Signal." J. Opt. Soc. Am., Vol. 59, No. 9, pp 1247–1248 (1969).
6. K. Krapels, R. G. Driggers, R. H. Vollmerhausen, N. S. Kopeika, and C. E. Halford, "Atmospheric turbulence modulation transfer function for infrared target acquisition modeling," Opt. Eng. **40**(9) 1906-1913 (2001).
7. D. Sadot, D. Shemtov, and N. S. Kopeika, "Theoretical adn experimental investigation of image quality through an inhomogeneous turbulent medium," Waves Random Media **4**(2), 177-189 (1994).

8. Resonators and Lasers

In this section, modeling of optical resonators is discussed. Resonators may be classified as either stable or unstable. Stable resonators were the first ones developed because they can be made to lase in low gain media. Stable resonators are distinguished by having a stable geometric mode. Typically these devices have low diffraction losses. As a consequence they are very slow to converge. Unstable resonators are not geometrically stable and generally have high losses. These devices are very useful for high gain media and have the important property that high beam quality is achievable.

8.1 Bare-Cavity Resonators

If we represent the starting complex amplitude distribution by ϕ_0, the complex amplitude distribution after one pass by φ_1, and the transformation operation $\mathbf{S}$; then the resonator operation is

$$\phi_1(x, y) = \mathbf{S}\phi_0(x, y). \tag{8.1}$$

For special ψ functions called eigenfunctions, we have the relationship

$$\mathbf{S}\psi(x, y) = \lambda\psi(x, y), \tag{8.2}$$

where λ is a scalar, complex coefficient called the eigenvalue. In the general case, there are an infinite number of eigenfunctions, each with its own eigenvalue. Both stable and unstable resonators are characterized by having eigenfunctions and eigenvalues which represent the round-trip loss coefficients. Many sophisticated methods of analysis have been developed which can find not only the lowest-loss modes but higher order modes as well. We state without proof that the eigenmodes are orthogonal and are a complete set. A combination of eigenmodes may be used to represent an arbitrary complex amplitude distribution in the resonator.

The lowest-loss mode is readily found with a physical optics code. The principle for finding these modes is simply that any arbitrary starting distribution may be fit to a set of resonator modes. Let $\phi_0(x, y)$ be the starting distribution, then

$$\phi_0(x, y) = \sum_{n=0}^{\infty} a_n \psi_n(x, y), \tag{8.3}$$

where a_n are the coefficients which result in the best fit of the eigenmode function set to the starting distribution. If c_n are the eigenvalues, then after m round trips the distribution in the resonator is

$$\phi_m(x, y) = \sum_{n=0}^{\infty} c_n^m a_n \psi_m(x, y) . \tag{8.4}$$

After a suitable number of passes, the mode with the largest eigenvalue will be the dominant term

$$\phi_m(x, y) \approx a_j c_j^m \psi_j(x, y) , \tag{8.5}$$

where it is assumed that ψ_j is the lowest-loss mode. In most cases this is the lowest order mode ψ_0. Provided the starting function has at least some finite amount of ψ_j, i.e., $a_j \neq 0$, the resonator will, in due course, converge to the lowest-loss mode. The selectivity is due to the apertures of the resonator. If one attempts to model a resonator numerically by setting up the mirrors with the proper focal lengths and spacings but does not include apertures, then all the eigenvalues will meet the condition $|c_n| = 1$ and the calculation will never converge. The diffraction loss at the apertures is what causes mode selectivity.

For modeling bare-cavity stable resonators it is convenient to include an energy renormalization at each cycle. For unstable resonators the round trip magnification is not unity. A rescale at the end of each cycle is necessary in addition to the energy renormalization to keep the matrix units of the problem constant. The gain medium will affect the transverse mode solution by nonlinear saturation effects and by the finite extent of its volume.

8.1.1 Flat-Flat Resonators

Perhaps the first resonator one might consider consists of two flat mirrors. The flat-flat resonator was prominent in the early laser investigations and is still employed from time to time. In fact, this configuration is beset with problems. Geometrically, it is easily understood that a ray which is at a slight axis to the mirror normals will “walk-off” the configuration. A similar walk-off occurs if the mirror are not perfectly parallel. Even in numerical analysis, where the alignment may be made ideal, flat-flat mirrors are prone to convergence difficulties.

Consider an example of a flat-flat resonator. A uniform beam is injected into the device and mode structure is observed during the 29 round trips needed to achieve convergence to 0.5%. The configuration is shown in Fig. 8.1 and the parameters are given in Table 8.1. The initial commands (as listed in Fig. 8.2) establish a 32 x 32 array, the wavelength at 10.6 microns, and a choice for the units (spacing between matrix points) of 0.15 cm. The beam is initially established as uniform intensity and a clear aperture is imposed. The energy in the beam is normalized to unit value. The command `geodata` is used to specify the surrogate gaussian beam properties to override the automatic determination.

Table. 8.1. Parameters or flat mirror resonator.

length	30 meters
aperture radius	1 cm
wavelength	1 μ

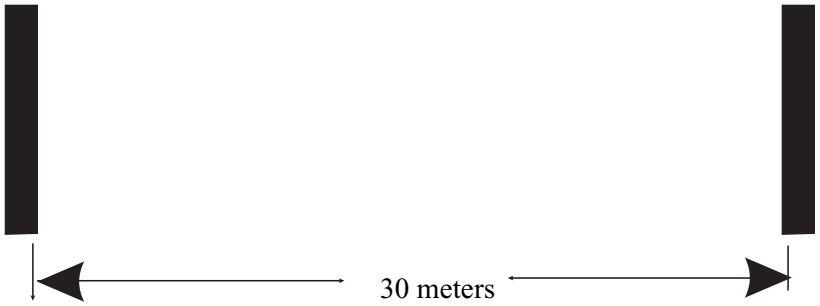


Fig. 8.1. Elementary resonator consisting of flat mirrors.

The cyclical behavior of the resonator is modeled by using a macro. The macro may be executed any number of times. A convergence test is used to exit from the macro if the energy is sufficiently converged. The integer register keeps track of the pass number. The real register is used for the x-coordinate of the summary plot made by `plot/udata`. The `prop` command does the propagation. `geodata` reestablishes the surrogate gaussian beam parameters — in effect overriding the automatic algorithm selection and insuring that each pass has exactly the same numerical treatment.

The energy at the end of a round trip is stored. A running average of the energy values is taken. When the running average is changing less than the amount established in gain/convergence/set (on the next page), the program exits from the macro. After the macro is defined, the macro is called for 40 executions. The round-trip loss is plotted in Fig. 8.3. The evolution of the transverse mode is given in Fig. 8.4.a. The device converges to 0.005 in 29 cycles. The transverse mode structure is changing appreciably during the first few round trips then goes through some strange gyrations before finally stabilizing to a near-gaussian form. After about 20 round trips, the transverse mode is well established to a near-gaussian mode. The relatively rapid convergence is not typical of stable devices. This example was deliberately designed with under-sized mirrors to assure rapid convergence.

8.1.2 Stable Resonators

The basic properties of resonators are determined from the first-order properties of the system. The ABCD matrix provides an especially convenient theoretical description of matrices. Siegman in *Lasers* presents an excellent discussion of eigenvalues and eigenmodes in Chaps. 15 and 16. The paraxial properties of the resonator determine the diffraction behavior. Siegman shows that the conditions of stability determined from diffraction theory using the Hermite-gaussian modes is exactly the same as for the paraxial behavior. In this section we shall only consider the paraxial ray behavior.

Let the first-order ray be characterized by

$$\boldsymbol{r} = \begin{bmatrix} y \\ \omega \end{bmatrix}, \tag{8.6}$$

```
c
c Flat mirror resonator
c
variab/dec/int pass STOP
array/s 1 32 # choose small array for fast execution
color 0 10.6 # set wavelengths
units 0 .15 # guess at units
clear 1 1 # start with uniform intensity
clap/c/c 1 1. # initial aperture
energy/norm 1 1 # normalize energy
c
c Force surrogate gaussian beam properties
c
c bean no. zwaistx zwaisty waistx waisty iplanex iplaney
 geodata 1 0 0 2 2 1 1 1 1
plot/w hmwka.plt
plot/l
title start
macro/def hmwk/o
 pass = pass + 1 # increment pass counter
 step = step + 1 # increment step number
 prop 3000
 mirror/flat
 clap/c/c 1 1. # 1.0 cm. radius aperture
 geodata 1 0 0 2 2 1 1 1 1 # reset gaussian beam
 prop 3000
 mirror/flat
 clap/c/c 1 1. # 1.0 cm. radius aperture
 geodata 1 0 0 2 2 1 1 1 1 # reset gaussian beam
 energy # calculate energy in the beams
 title Mirror 2
 plot/d hmwk@pass.plt/o # off-line plot
 plot/l
 plot/w hmwkb.plt/o # off-line plot
 plot/l
 variab/set/param Energy 1 energy
 Energy = Energy - 1 # calculate energy difference
 udata/set pass step Energy # store energy differences
 plot/w hmwka.plt
 title Energy loss per pass
 plot/udata first=1 last=1
 gain/converge/test 1 STOP # test for convergence
 pause 2 # pause for 2 seconds to see data
 energy/norm 1 1 # renormalize energy
 if STOP macro/exit # exit from macro on convergence
macro/end
pass = 0 ; step = 0 # initialize variables
gain/converge/set eps1=.005 3 # set convergence criterion to
```

Fig. 8.2. Input commands for flat-flat mirror resonator.

where y is the ray height above the axis and ω is the optical angle. The ABCD matrix may be represented as the system matrix $\mathbf{M}$,

$$\mathbf{M} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}. \tag{8.7}$$

The eigenequation is

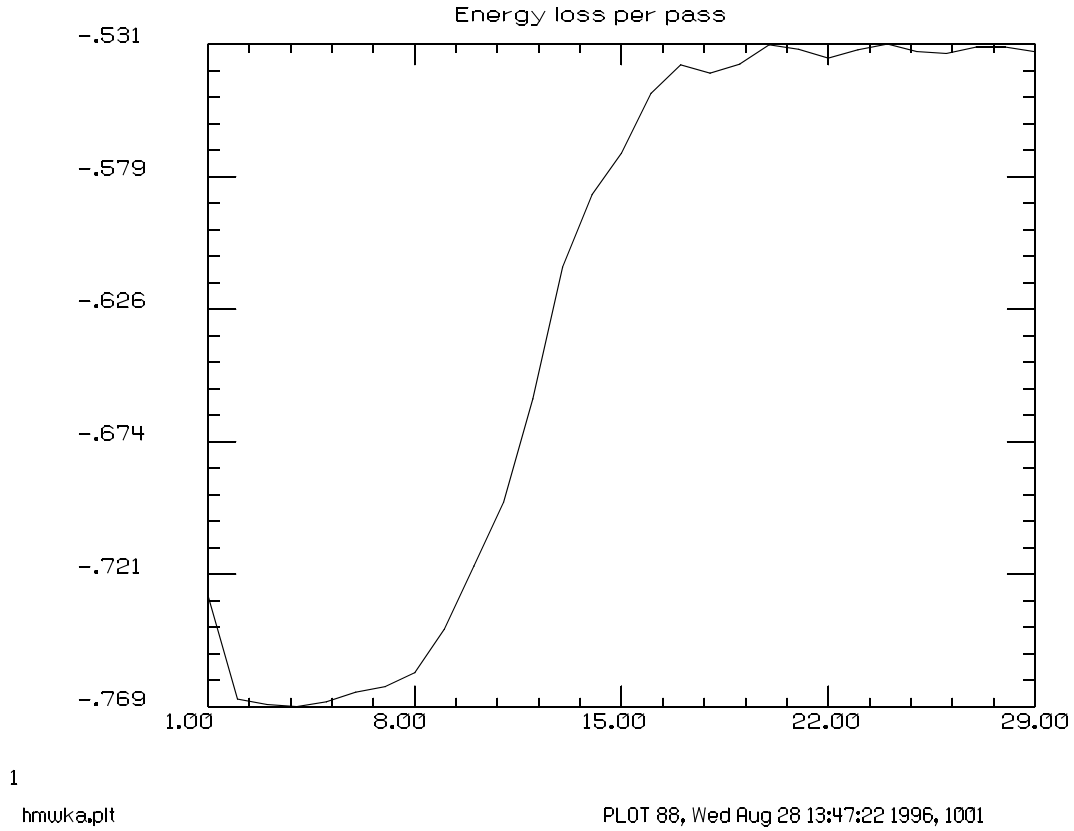


Fig. 8.3. Convergence after about 29 round trips.

$$\mathbf{Mr} = \lambda \boldsymbol{r}. \tag{8.8}$$

This is equivalent to the expression,

$$(\mathbf{M} - \lambda \mathbf{I})\boldsymbol{r} = 0, \tag{8.9}$$

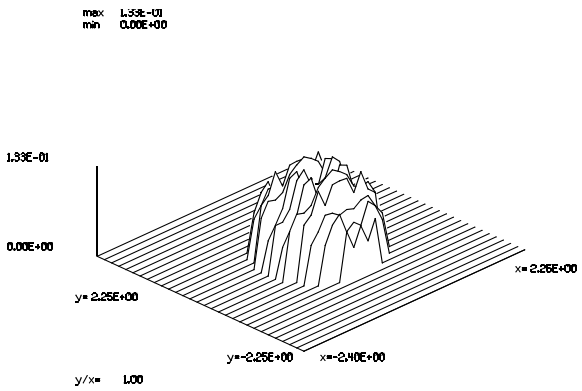
where $\mathbf{I}$ is the identity matrix. Non-trivial solutions of Eq. (8.9), i.e., those with $\boldsymbol{r} \neq 0$, occur if and only if $\det[\mathbf{M} - \lambda \mathbf{I}] = 0$,

$$\begin{vmatrix} A-\lambda & B \\ C & D-\lambda \end{vmatrix} = \lambda^2 - (A+D)\lambda + 1 = 0. \tag{8.10}$$

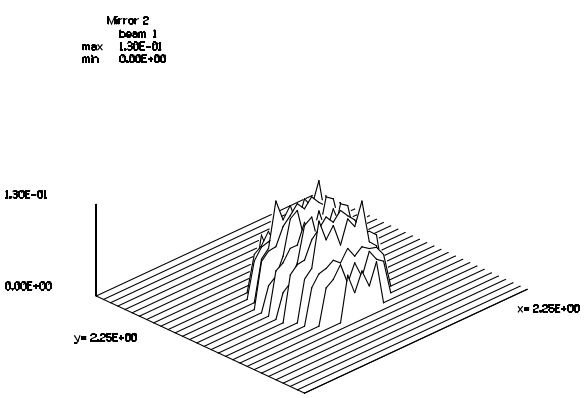
Taking note that $AD - BC = 1$. We define the m-parameter as,

$$m \equiv \frac{A+D}{2}. \tag{8.11}$$

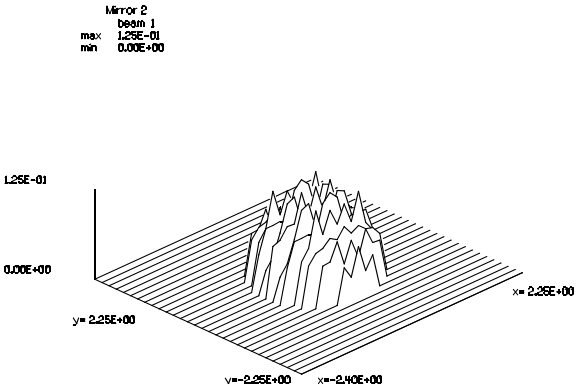
The roots of Eq. (8.10) are



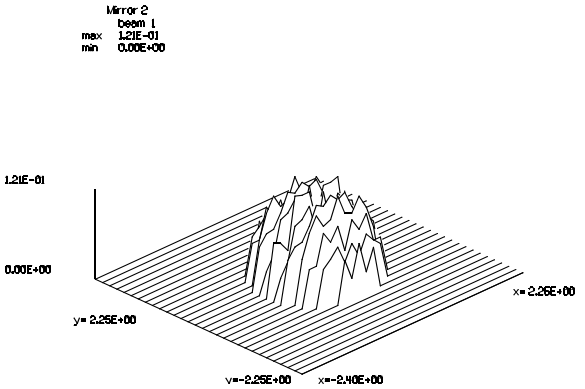
After Pass 1



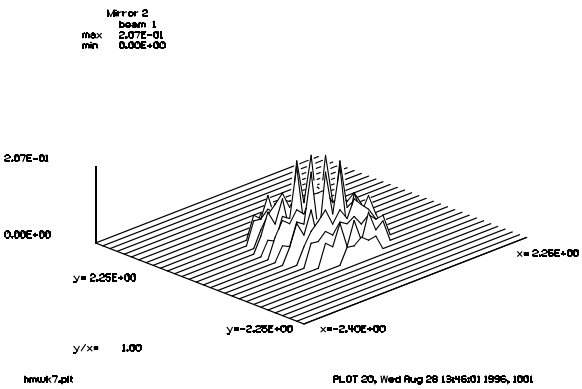
After Pass 2



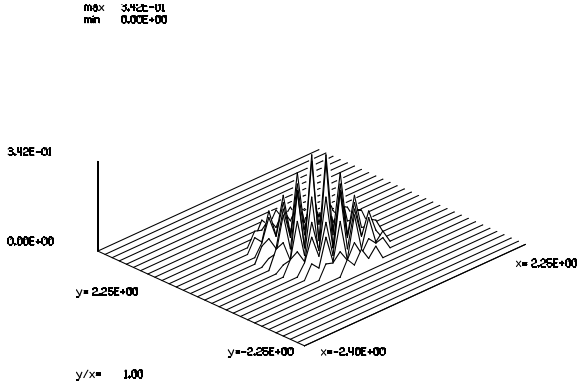
After Pass 3



After Pass 5



After Pass 7

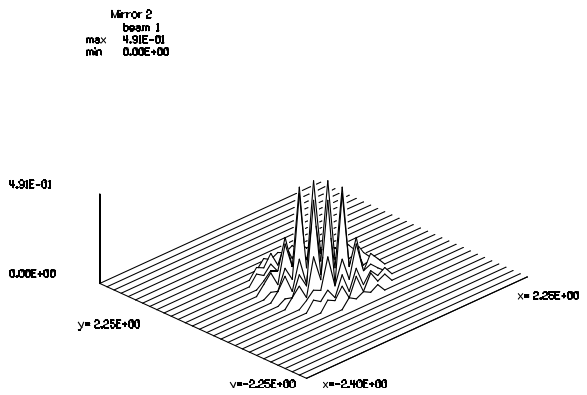


After Pass 9

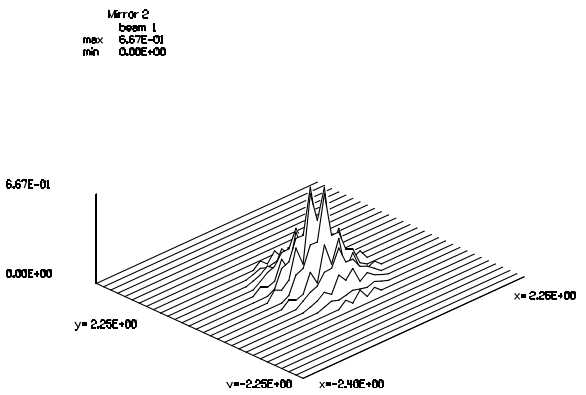
Fig. 8.4.a. Transverse mode after various passes through the flat-flat resonator of Fig. 8.1. Note that passes 1 through 5 are similar and that a mode transition occurs during passes 7 through 9.

$$\lambda_a, \lambda_b = m \pm \sqrt{m^2 - 1}\,. \tag{8.12}$$

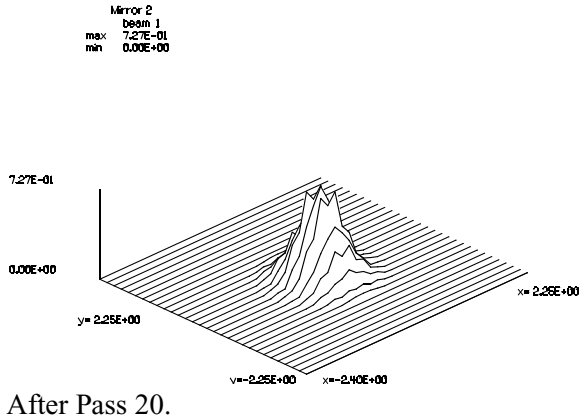
We note that $\lambda_a \lambda_b = 1$. More specifically, if



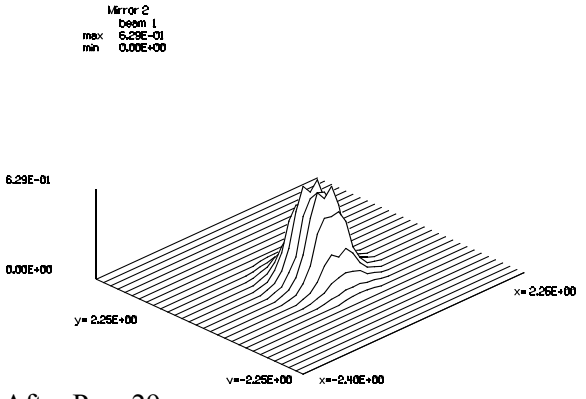
After Pass 11.



After Pass 13.



After Pass 20.



After Pass 29;

Fig. 8.4b. Transverse mode after various passes through the flat-flat resonator of Fig. 8.1. The transition begun at about pass 5 stabilizes at about pass 20 to a near-gaussian form.

$$|m| \leq 1\,, \tag{8.13}$$

the system is stable and we can write Eq. (8.12) as,

$$\lambda_a, \lambda_b \;=\; m \pm j\sqrt{1-m^2} \;=\; \cos\theta \pm j\sin\theta \;=\; e^{\pm j\theta}\,. \tag{8.14}$$

The eigenvalues are unit value phasors and are mutually conjugate.

Because the eigenvectors are, in general, not strictly real, there are no single pass eigensolutions for real paraxial rays. If there exists N such that $N\theta = \pi$, then there are real paraxial ray eigensolutions for N systems joined together to form a composite system identified as $\mathbf{M}_N$ such that in relation to the elementary system it has the property that $\mathbf{M}_N = \mathbf{M}^N$. Such a real ray is then an eigensolution of a system consisting of N elementary systems.

The eigenvectors may be found by forming the modal matrix which is

$$\text{modal matrix} = \text{adj}[\mathbf{M} - \lambda\mathbf{I}]_{\lambda_a, \lambda_b} = \begin{vmatrix} D - \lambda_a & -B \\ -C & A - \lambda_b \end{vmatrix}. \tag{8.15}$$

The eigenvectors are therefore,

$$\boldsymbol{r}_a = \begin{bmatrix} D - \lambda_a \\ -C \end{bmatrix} \text{ and } \boldsymbol{r}_b = \begin{bmatrix} -B \\ A - \lambda_b \end{bmatrix}. \tag{8.16}$$

An arbitrary input may represented as a sum of eigenfunctions,

$$\phi = \sum_{n=1}^{\infty} a_n \psi_n . \tag{8.17}$$

For the specific case of geometrical rays where there are just two solutions. An arbitrary initial ray vector $\boldsymbol{r}_0$ may be written in terms of the two eigenvectors

$$\boldsymbol{r}_0 = c_a \boldsymbol{r}_a + c_b \boldsymbol{r}_b . \tag{8.18}$$

After n cycles the ray vector is,

$$\boldsymbol{r}_n = c_a \boldsymbol{r}_a e^{jn\theta} + c_b \boldsymbol{r}_b e^{-jn\theta} = \boldsymbol{r}_0 \cos n\theta + j\boldsymbol{s}_0 \sin n\theta , \tag{8.19}$$

where $\boldsymbol{s}_0 = c_a \boldsymbol{r}_a - c_b \boldsymbol{r}_b$, which may be considered the input slope vector.

Equation (8.14) indicates oscillatory behavior. The m-parameter,

$$m \equiv \frac{A + D}{2} = \cos\theta , \tag{8.20}$$

guarantees stable performance if $|m| < 1$. The oscillatory behavior vanishes for $\theta = 0$ and $\theta = \pi$ — at the limits of stability.

The treatment above allows analysis of very general devices. There is, in principle, no limit to the complexity of the system. If the round trip ABCD system matrix is known, the stability can be determined. A less general treatment is often given which uses the g-parameters

$$g_1 = 1 - \frac{L}{R_1}, \; g_2 = 1 - \frac{L}{R_2}, \tag{8.21}$$

where the sign convention traditional in resonator theory is that R is positive if the mirrors are concave-inward and Mirror 1 is on the left and Mirror 2 is on the right, as illustrated in Fig. 8.5.

Using the g-parameters, the stability criterion is

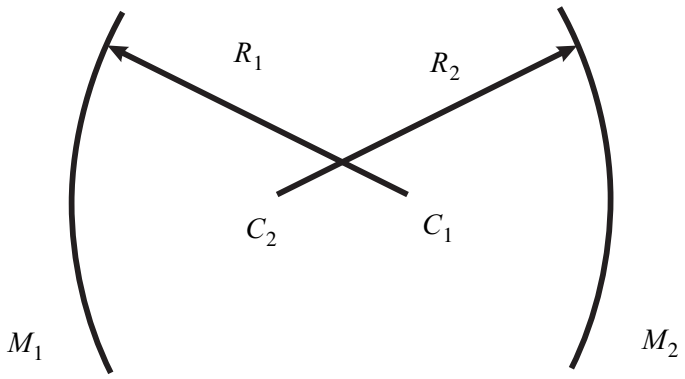


Fig. 8.5. Elementary two-mirror resonator. The two mirrors shown have positive radii according to the special sign convention used for resonators.

$$0 < g_1 g_2 < 1 \,. \tag{8.22}$$

While the upper limit of stability for the g-parameter is the same as for the m-parameter at unity (see Eq. 8.13, the numeric value of $g_1 g_2$ is not the same as m within the stable regime. The m-parameter has the advantage that it may be applied to any system no matter how complex, whereas the g-parameter is restricted to the two-mirror case.

For a symmetrical system consisting of two concave mirrors, we have stability if the centers of curvature overlap, i.e., if c_2 is to the left of c_1, where c_1 and c_2 are the centers of curvature. We can, therefore, tell at a glance whether a two-mirror system is stable.

A ray, in a stable system, will oscillate forever, provided the mirror apertures are sufficiently large. If we adjust the system so that $|m| \to 1$, then $\theta \to 0$ and the rate of oscillation goes to zero. These conditions arise, for example for either planar mirrors or concentric mirrors. Both are at the stability limit. For plane mirrors, the points c_1 and c_2 approach ∞ and $-\infty$ respectively. At infinity, these points are identical and there in no overlap, i.e., the stability criterion is not met. Slight misalignment of the plane mirrors causes the ray to walk off sideways. Slight misalignment of the concentric system causes the rays to rotate until the ray misses the apertures.

As $|m| \to 0$, $\theta \to \pi/2$ which occurs along both the vertical and horizontal axes, one of the mirrors lies at the center of curvature and the oscillation will be at a maximum rate. Consider a flat mirror at the center of curvature of the other mirror. The system is completely insensitive to rotation of the flat mirror, but axial translation of the flat mirror can cause instability.

The regions of stability and unstability can be mapped into g-parameter space, as shown in Fig. 8.6. A resonator is stable if it lies between the horizontal or vertical axes and the two sheets of the hyperbola defined by $g_1 g_2 = 1$ and the horizontal and vertical lines for $g_1 = 0$ and $g_2 = 0$.

Three special conditions are of particular interest: two plane parallel mirrors, confocal symmetrical, and concentric. The later two types are marginally stable.

We can represent the gaussian beam in terms of the complex q-parameter,

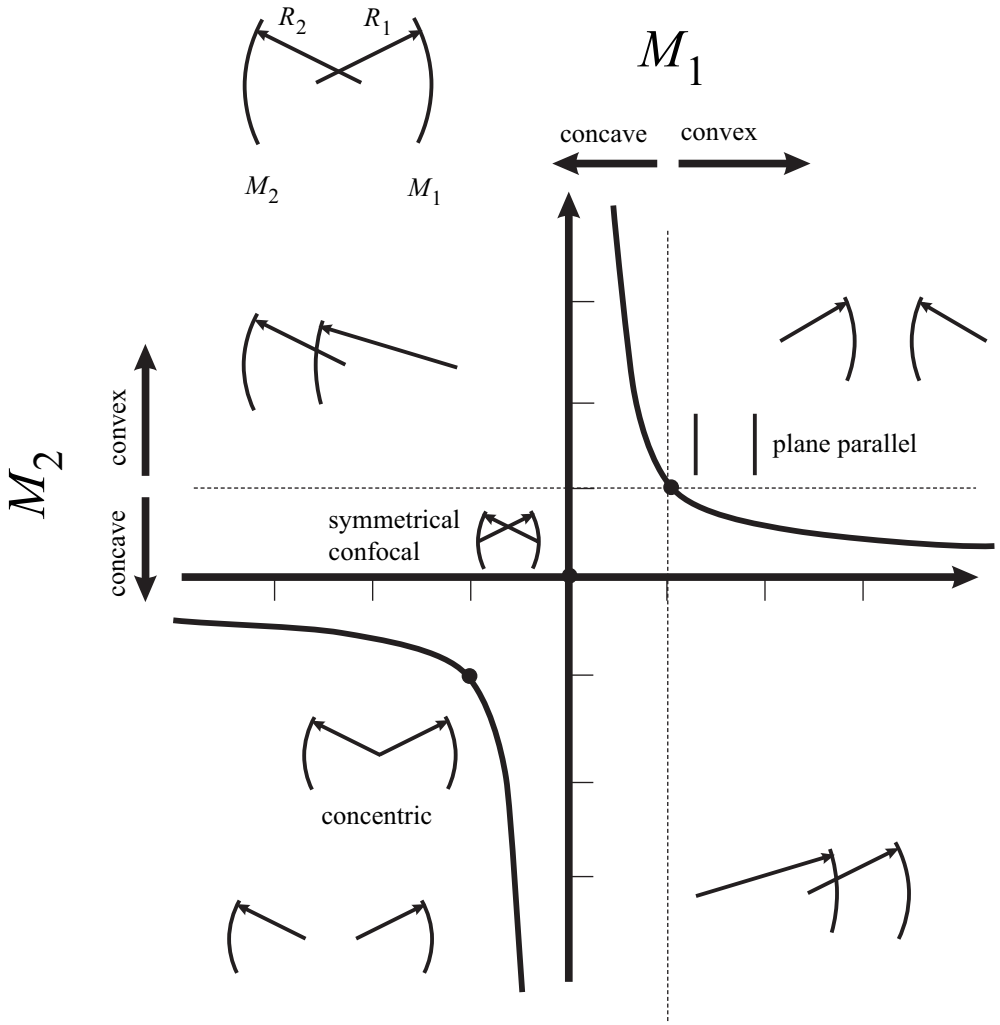


Fig. 8.6. Stability diagram for g_1 and g_2 The region between the axes and the two sections of the hyperbolic curves is stable. All other points are unstable.

$$\frac{1}{\tilde{q}} = \frac{1}{R} - \frac{j\lambda}{\pi\omega^2}. \tag{8.23}$$

The complex q-parameter may be considered to be the complex phase curvature. From first-order optics we have the relationship,

$$\tilde{R}_2 = \frac{A\tilde{R}_1 + B}{C\tilde{R}_1 + D}, \tag{8.24}$$

where the tilde indicates the reduced radius, radius divided by the index of refraction. A similar equation holds for the q-parameter — a form of complex inverse radius.

$$\tilde{q}_2 = \frac{A\tilde{q}_1 + B}{C\tilde{q}_1 + D}. \tag{8.25}$$

To find the gaussian eigenmode solution, we solve the equation,

$$\tilde{q} = \frac{A\tilde{q} + B}{C\tilde{q} + D}, \tag{8.26}$$

with the solutions,

$$\frac{1}{\tilde{q}} = \frac{D-A}{2B} \pm \frac{1}{B}\sqrt{\left(\frac{D-A}{2}\right)^2 + BC}. \tag{8.27}$$

Given the eigenmode q-parameter, the local transverse radius and phasor radius are determined by,

$$R = \frac{2B}{D-A}, \quad \omega = \left(\frac{\lambda}{\pi}\right)^{1/2} \frac{|B|^{1/2}}{\left[1 - \left(\frac{A+D}{2}\right)^2\right]^{1/4}}, \tag{8.28}$$

and the waist properties determined by,

$$z_{\text{waist}} = \frac{R}{1 + \left(\frac{\lambda R}{\pi \omega^2}\right)^2}, \quad \omega_0 = \frac{\omega}{\sqrt{1 + \left(\frac{\pi \omega^2}{\lambda R}\right)^2}}. \tag{8.29}$$

First, we note that Eq. (8.27) has the same condition for existence as the paraxial ray solution for stability. We choose the root that gives a negative imaginary part — the other root leads to a solution which does not fall off away from the axis.

Once we have the gaussian eigenfunction, we have solved for the lowest-loss gaussian mode. For many stable resonator configurations, this elementary analysis gives an accurate answer. The calculation is very easy to implement on a spread sheet program. In designing and analyzing a stable resonator my first step is to build a spread sheet model of the problem to verify stability and to establish the locations of waists and the size of the beam at various points.

It is important to consider that the gaussian mode was determined using only the optical powers and spacings. The aperture size played no role. Such an idealized device would have no losses and therefore no mode selectivity. We could get an approximate idea of the losses by considering the relative size of the gaussian eigenmode at each aperture and determining the magnitude of the clipping.

It is relatively easy to test a numerical configuration for stability by propagating a gaussian beam. We can create a gaussian beam with waist at the proper position and correct beam radius. No apertures are used in this numerical check. The beam is propagated through several round trips and the beam size is checked at each point. If the beam size is not exactly reproduced, the resonator can be adjusted by varying the distance or radius of one of the mirrors.

The first example is a stable resonator with circular mirrors, as shown in Fig. 8.7.a. The configuration consists of a flat mirror and a concave spherical mirror of radius 50 cm. The mirrors are separated by 45 cm. The parameters are summarized in Table 8.2.

Table. 8.2. Parameters for stable resonator example.

length	45 cm
mirror radius	50 cm
wavelength	1.064 μ
Rayleigh range	15 cm
waist radius	0.02253936 cm
aperture radius	0.14 cm

It is important to establish a geometrically stable resonator before attempting to calculate the diffraction performance. A surrogate gaussian beam is used to determine the switching of the propagation algorithms. If the surrogate gaussian is not set to the geometric resonant condition, the parameters will oscillate without damping. In principal the diffractive performance would be unaffected. Because of sampling limitations an oscillation in the selection of propagation algorithms will perturb the diffractive mode significantly — preventing convergence.

Under these conditions, an actual distribution may be used as a seed to find the diffractive transverse mode as determined by the apertures and aberrations, if any. In this example, Beam 1 is started with the ideal geometric mode as a seed and Beam 2 is started with a uniform intensity distribution. Beam 1 converges almost immediately to 0.09 percent loss per pass for the aperture radius of 0.14 cm. Beam 2 takes about 90 passes to converge to within 0.1 percent. In this example two beams were used simultaneously. The first mode was the ideal geometric mode which should be very close to the true diffraction mode. The second mode is a simple plane wave. Fig. 8.7b plots the energy as a function of cycle number for the two modes. The initial gaussian beam converges almost immediately to a stable, low loss value. The plane wave takes about 90 cycles to converge to an acceptable degree. The ideal resonator mode is shown in Fig. 8.8.a and the mode determined by numerical analysis is shown in Fig. 8.8b. The modes are indistinguishable visually.

If the resonator configuration is significantly modified, the simple gaussian mode may no longer be the lowest-loss mode. Consider the resonator from the previous example with the addition of an obscuration at the center of the flat mirror, as shown in Fig. 8.9. If this obscuration is very small, the mode will be largely unaffected but if the hole is increased sufficiently, a mode which has low intensity in the center may prevail in the mode competition. The parameters of the resonator are listed in Table 8.3.

The aperture size has been reduced to make convergence faster. In the numerical experiment, the obscuration is increased in steps of 0.005 cm for ten steps and the resonator is allowed to settle for 25 round trips. Where there is a clearly defined mode, the laser stabilizes easily in the 25 passes but where two modes have nearly the same round-trip loss, 25 passes are insufficient for settling.

The round-trip energy loss is plotted against the number of passes in Fig. 8.10. At the first set of 25 passes, the obscuration is so small that only the center point is obscured. After the second significant increase in obscuration — at 50 passes — there is significant disruption of the settling, indicating modes

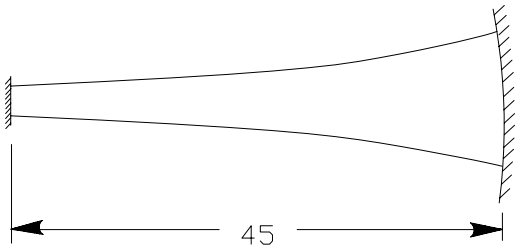


Fig. 8.7.a. Simple resonator consisting of a flat mirror and a concave mirror. The waist will form at the flat mirror.

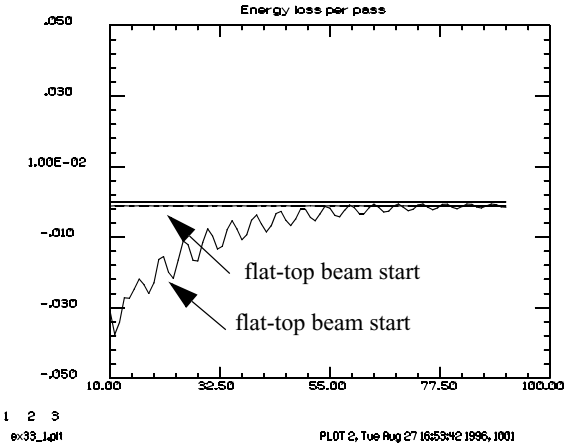


Fig. 8.7b. Plot of energy loss per pass as a function of the pass number. The plot is from Pass 10 to Pass 91. The lower horizontal line is actually the loss for Beam 1 which has already converged from the geometric mode within 10 passes. The oscillating curve shows the convergence of Beam 2 from the initial uniform intensity.

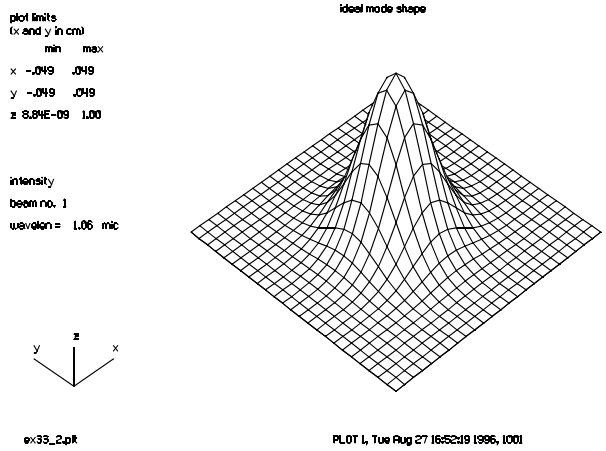


Fig. 8.8.a. Ideal resonator mode.

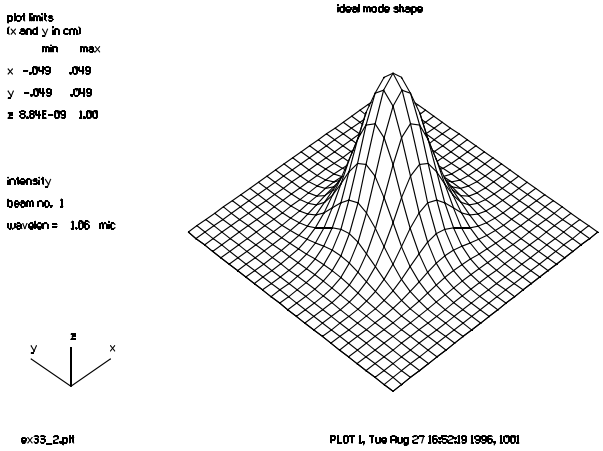


Fig. 8.8b. Converged numerical mode. The ideal and calculated modes are virtually identical.

keeping. Examination of the isometric plots on the following pages can be compared with the regions of this plot to see the mode changes.

The resonator stayed in the lowest order gaussian mode until about the 100th pass — after the obscuration had been increased three times. We see the formation of a different mode after 150 passes. At higher obscuration values, we start to get significant azimuthal variation. The finite accuracy of the numerical calculations plays an increasing role as the obscuration increases because so many modes have similar eigenvalues. It is interesting to note that the azimuthal direction tends to break up in to lumps of about the size of the Airy pattern.

The transverse modes of ideal stable resonators take the form of Hermite gaussian polynomials or Laguerre gaussian. The general polynomial form of the Hermite-gaussian functions is

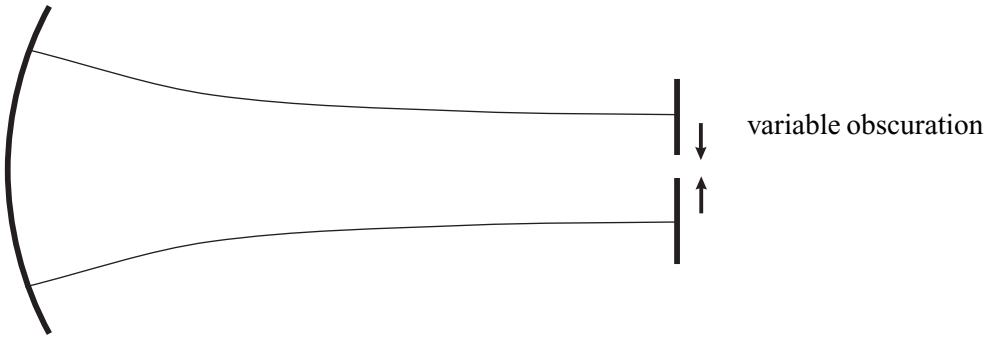


Fig. 8.9. Concave and flat mirror resonator with variable size hole in the flat which forms an obscuration in the beam.

Table. 8.3. Stable resonator example with variable aperture.

length	45 cm
mirror radius	50 cm
wavelength	1.064 μ
Rayleigh range	15 cm
waist radius	0.02253936 cm
aperture radius at concave mirror	0.12 cm
obscuration radius at flat mirror	(variable)

$$u_n(x) = \left(\frac{2}{\pi}\right)^{1/4}\left(\frac{1}{2^n n!\,\omega_o}\right)^{1/2} H_n\left(\frac{\sqrt{2}x}{\omega(z)}\right)e^{-x^2/\omega^2} \tag{8.30}$$

where n is the order of the polynomial, ω_0 is a waist radius parameter similar to the gaussian beam and $H_n(x)$ are the Hermite functions. The two-dimensional functions may be described by multiplying two one-dimensional functions. The order and waist parameter may be different for the two directions. As an example, Hermite-gaussian mode (2,1) is illustrated in Figs. 8.11.a and 8.11b.

The Laguerre-gaussian functions, defined at the waist, are described in cylindrical coordinates,

$$u(r,\theta) = \sqrt{\frac{2p!}{(1+\delta_{0m})\pi(m+p)!}}\frac{e^{j(2p+m+1)}}{\omega_0}\left(\frac{\sqrt{2}r}{\omega_0}\right)^m L_p^m\left(\frac{2r^2}{\omega_o^2}\right)e^{-2r^2/\omega_0^2}e^{jm\theta}, \tag{8.31}$$

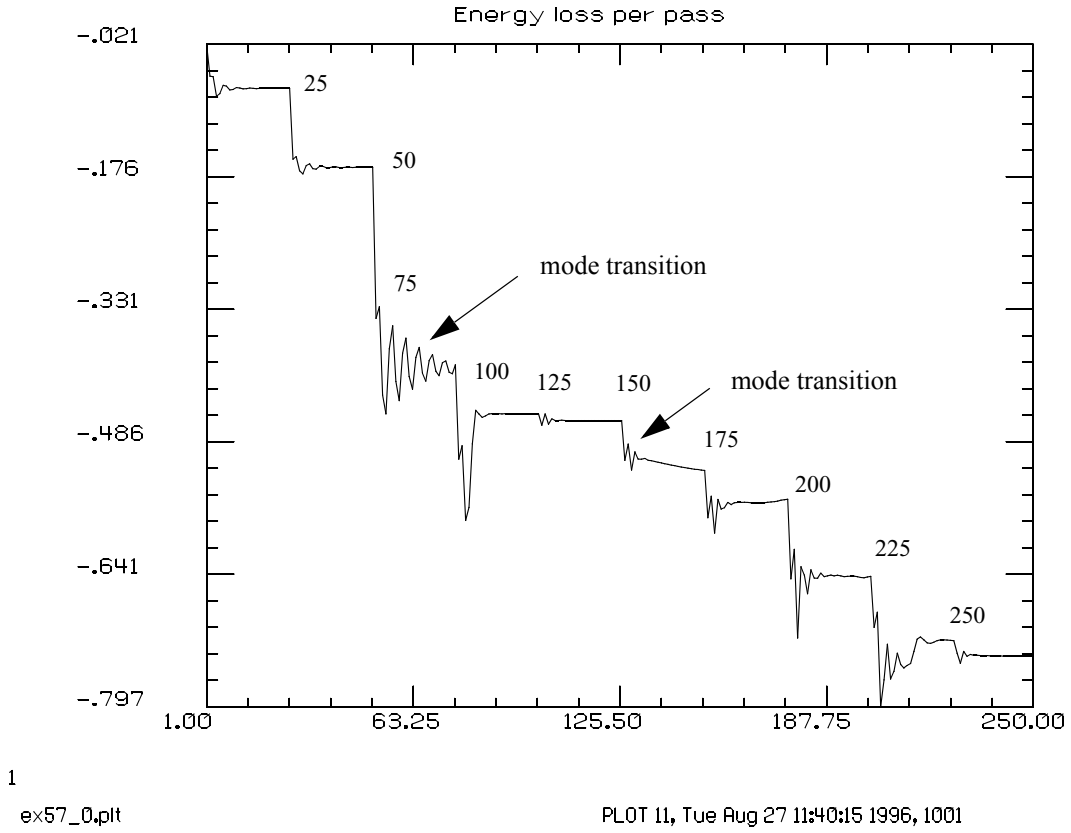


Fig. 8.10. Energy loss as a function of number of passes. Note the first region of poor convergence after 75passes.

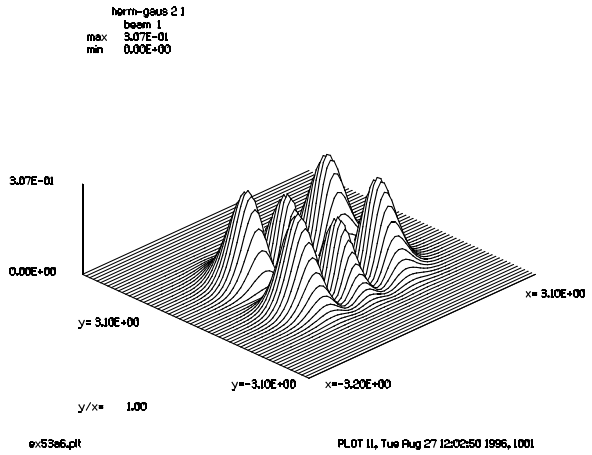


Fig. 8.11.a. Hermite gaussian mode 2,1.

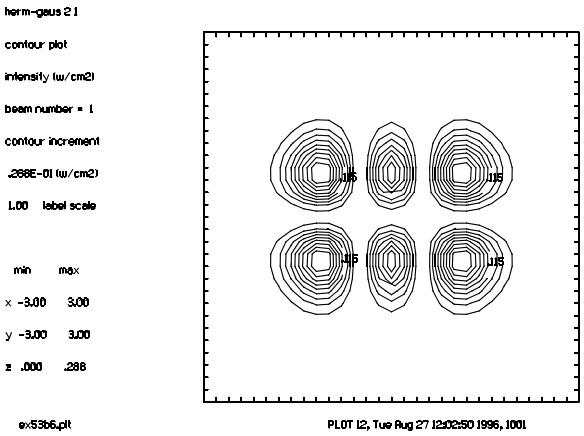


Fig. 8.11b. Hermite gaussian mode 2,1.

where θ is the azimuthal variation, p is the radial order, m is the azimuthal order, δ_{0m} takes the value 1 if $m = 0$ and zero otherwise, and L_p^m is the Laguerre polynomial. As an example, Laguerre-gaussian mode (2,1) is illustrated in Fig. 8.12.a and 8.12b.

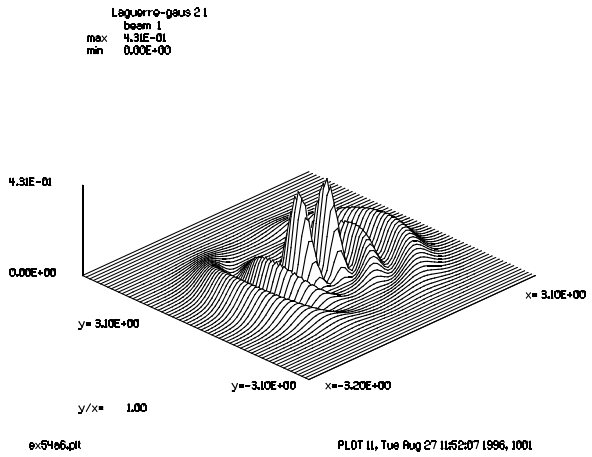


Fig. 8.12.a. Laguerre-gaussian mode 2,1.

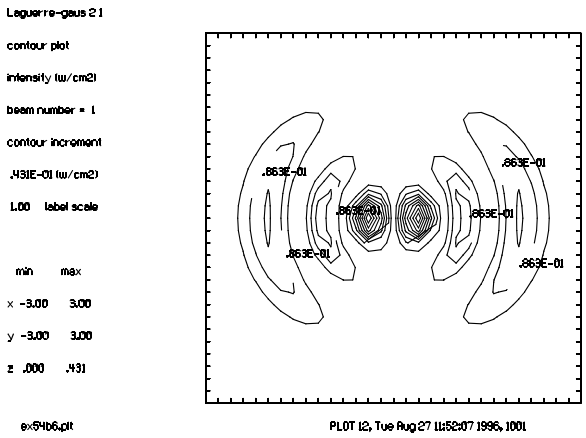


Fig. 8.12b. Laguerre-gaussian mode 2,1.

8.1.3 Aligned Unstable Resonator

The condition for a laser to be unstable is

$$|m| = \left|\frac{A+D}{2}\right| > 1\,. \tag{8.32}$$

Unstable resonators are divided into two general classes. If $m > 1$, the resonator is a positive branch unstable resonator and must have an even number of internal foci — most frequently zero internal foci. If $m < -1$, the resonator is a negative branch unstable resonator and has an odd number of internal foci. Internal foci present obvious problems for high power lasers.

We may find eigenvalues for the unstable resonator, as given in Eq. (8.33).

$$\lambda_a,\lambda_b \;=\; m \pm \sqrt{m^2-1} \;=\; M \text{ or } \frac{1}{M}, \tag{8.33}$$

where M is the round-trip magnification.

In general, we must find the eigenfunctions by numerical analysis. We can, however, find a pair of radii which are geometrical solutions to the eigenequation,

$$\frac{1}{R_a} \;=\; \frac{D-\lambda_a}{B}, \frac{1}{R_b} \;=\; \frac{D-\lambda_b}{B}\,. \tag{8.34}$$

In most cases we choose the solution with magnification greater than 1, called the “magnifying” solution rather than the “demagnifying” solution. The dehumidifying solution will collapse into the optical axis and ultimately, because of diffraction, will come back out as a magnifying solution, so it is usually sufficient to analyze the magnifying wave. Fig. 8.13 illustrates the ray paths of a typical unstable resonator.

Consider a confocal unstable resonator with circular mirrors. The configuration is the same as one described in Siegman and Miller [1]. The collimated and equivalent Fresnel numbers are

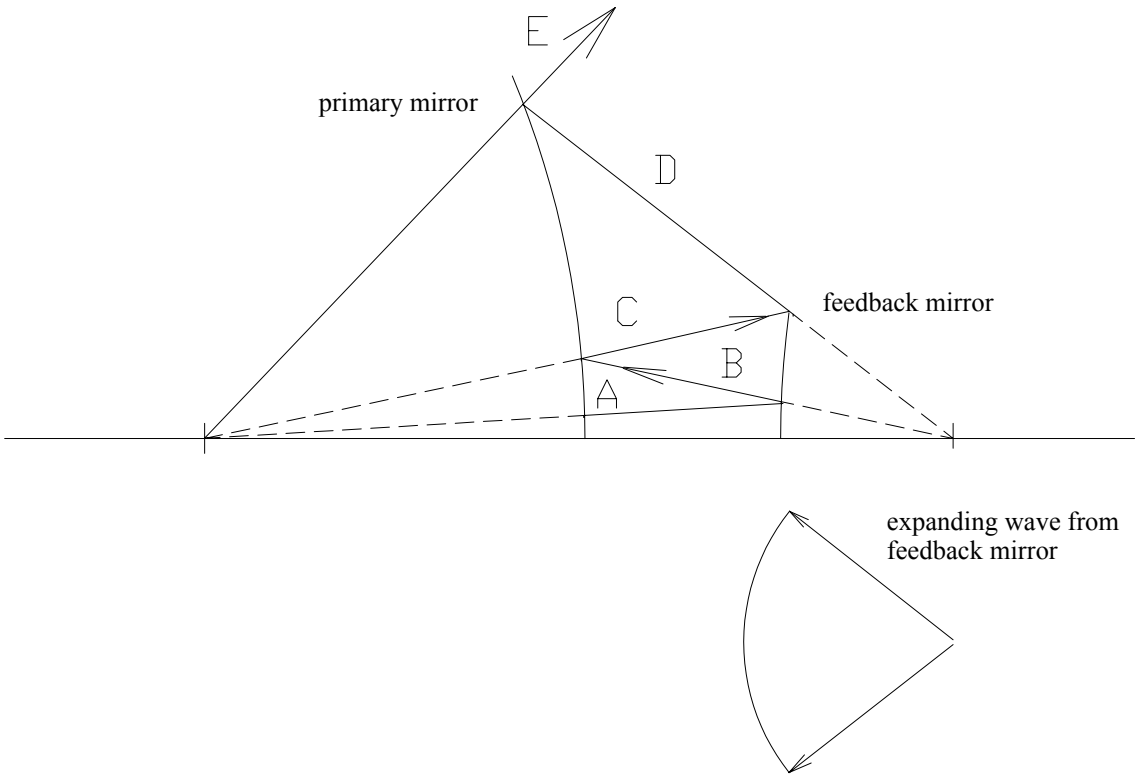


Fig. 8.13. Ray picture of unstable resonator. The ray goes through segments A through E, increasing in magnification by about 3 each time. The ray shown is set to originate from the center of curvature of the eigenradius. In this example, the inner third of the outgoing wavefront is fed back during each pass. In numerical analysis, the units will expand by the magnification in one round trip. After aperturing by the scraper mirror, leaving only the inner part of the array, we rescale the array to the original units, discarding the outer parts of the distribution, which are, of course, zero after aperturing.

$$N_c = \frac{Ma^2}{B\lambda}, N_{\text{eq}} = \frac{M^2 - 1}{2M}\frac{a^2}{B\lambda}, \tag{8.35}$$

where a is the radius of the secondary mirror, B is distance parameter from the round-trip ABCD matrix, and λ is the wavelength, and M is the magnification.

The parameters that are used are, $L = 90$ cm, $a = 0.3$ cm, and $\lambda = 0.001$ cm. This results in $N_c = 2$ and $N_{\text{eq}} = 0.75$. The configuration is shown schematically in Fig. 8.14.

After one round trip, the units of the distribution are twice those of the starting distribution. To start the next round trip the distribution must be rescaled to the original units. According to Siegman and Miller, the loss per cycle should be 44%. The numerical calculations are in close agreement with the published results. The resonator takes about 10 passes to converge. A plot of the energy loss per cycle is shown in Fig. 8.15.

The distribution after the scraper mirror is shown in Fig. 8.16.a. The obscured part of the beam is reflected back into the resonator to form the feedback. The far-field pattern is shown in Fig. 8.16b. The device is nearly diffraction limited.

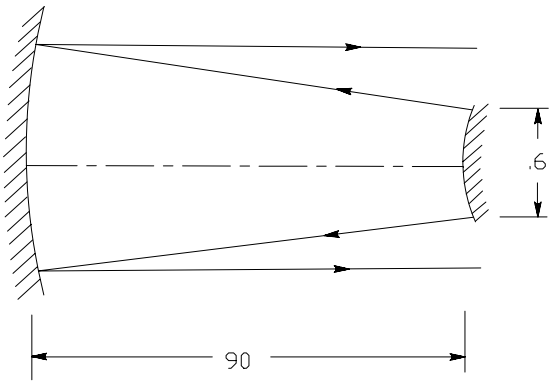


Fig. 8.14. Confocal unstable resonator with magnification of 2 per round trip.

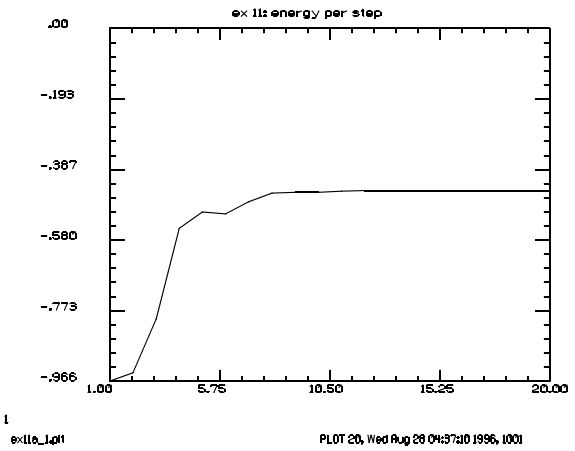


Fig. 8.15. Plot of energy loss per pass as a function of the pass number.

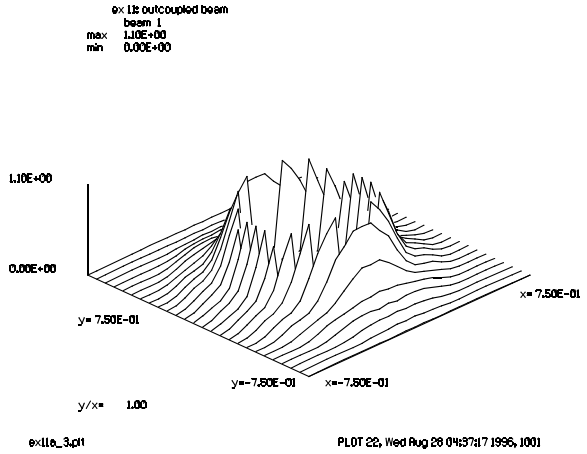


Fig. 8.16.a. Converged transverse mode after scraper mirror.

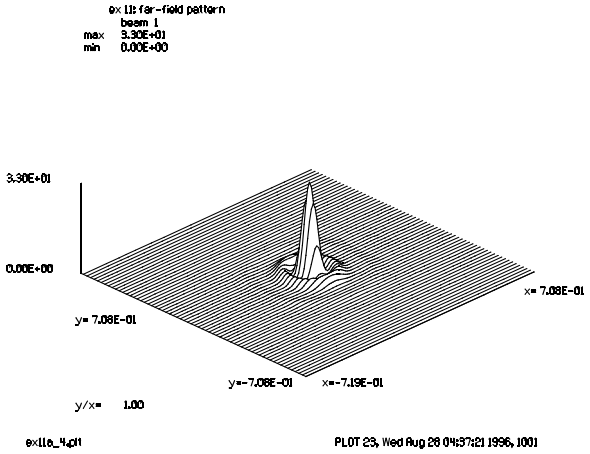


Fig. 8.16b. Far-field distribution of aligned resonator.

8.1.4 Misaligned Unstable Resonator

The next example is the same confocal, unstable resonator with the addition of 0.1 wave of tilt aberration representing misalignment. The loss per pass increases to 55 percent and the convergence takes longer. The misaligned resonator takes about 23 passes to converge.

A plot of the energy loss per cycle is shown in Fig. 8.17.a. The converged eigenmode, determined from numerical analysis in Example 12, is shown in Fig. 8.17b. The distribution after the scraper mirror is shown in Fig. 8.17c. The obscured part of the beam is reflected back into the resonator to form the feedback. The far-field pattern of the misaligned resonator is shown in Fig. 8.17d. Both the near- and far-field distributions are severely disturbed by the tilt of 0.1 wavelength.

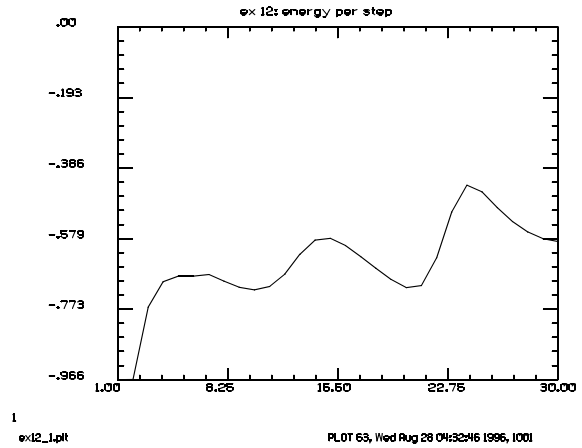


Fig. 8.17.a. Plot of energy loss per pass as a function of the pass number for misaligned resonator.

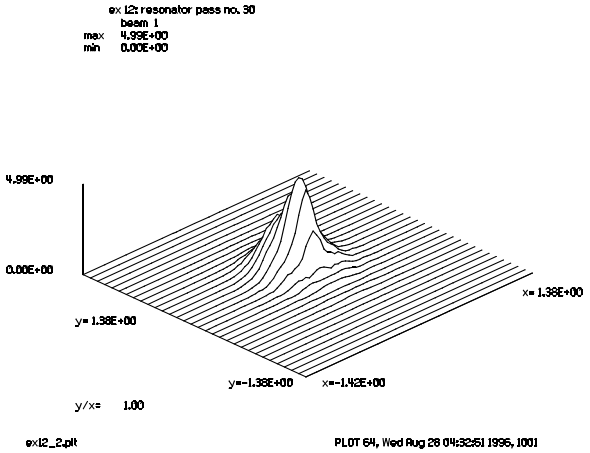


Fig. 8.17b. Converged transverse mode for misaligned resonator.

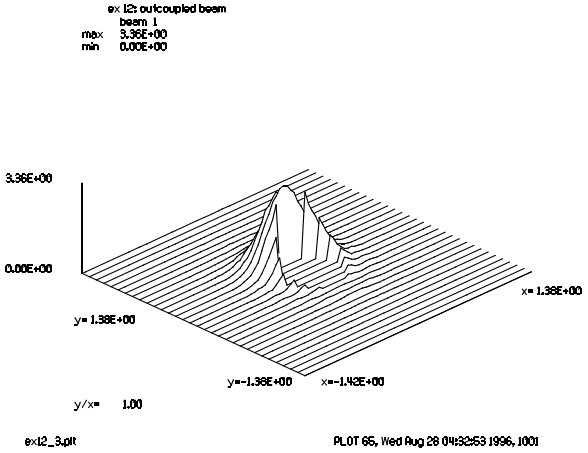


Fig. 8.17c. Converged transverse mode after scraper mirror for misaligned resonator.

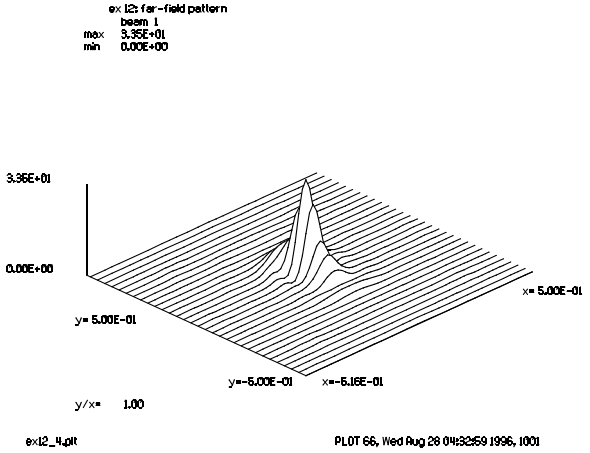


Fig. 8.17d. Far-field pattern of output of misaligned unstable resonator.

References

1. A. E. Siegman and H. Y. Miller, “Unstable Optical Resonator Loss Calculations Using Prony Method,” Appl. Opt. Vol. 9, No. 12, p. 2729 (1970).

9. Gain and Nonlinear Media

Propagation through active media involves both diffraction and gain or absorption. The numerical approach to solution is the split-step method described in previous chapters. All of the gain commands use the gain sheet approach and contain both gain and diffraction steps. In this chapter, the gain part of the inhomogeneous wave equation is described.

In general, gain is described as a function of the density of the active medium and the intensity of the optical field. Medium density influences the small signal gain and, in general, has some spatial variation. Because of saturation of the medium, the gain is a nonlinear function of the intensity of the optical field. A simple model of gain using Beer's Law (with a saturation intensity) may be used. The saturated form of Beer's Law may be represented by

$$I(z+\Delta z) = I(z)\exp\left(\frac{g_0\Delta z}{1+\frac{I(z)}{I_{sat}}}\right)^q, \tag{9.1}$$

where g_0 is the small signal gain, I_{sst} is the saturation intensity, and $q = 1/2$ for inhomogeneously and $q = 1$ for homogeneously broadened gain.

The gain grows exponentially at low values,

$$\frac{dI}{dz} \approx g_0 I. \tag{9.2}$$

The characteristic gain length is $1/g_0$. When I is comparable to I_{sat}, the homogeneously broadened gain takes the form

$$\frac{dI}{dz} \approx g_{sat} I_{sat}, \tag{9.3}$$

which is a linear increase in intensity. Figure 9.1 shows a typical case of homogeneously broadened gain. In the initial stages of amplification, the gain is exponential and in the later stages it is linearly increasing. In the saturated regime, the transverse mode may be modified slightly by the saturation of the mode so that the performance is not identical to the bare-cavity performance. Consider Figs. 3.4a and 3.4b which show diffraction patterns with and without saturation.

9.1 Beer's Law Gain

A simple model of homogeneously broadened gain using Beer's Law (with a saturation intensity) may be called with the `beer` command. The saturated form of Beer's Law may be represented by Eq. (9.1). The

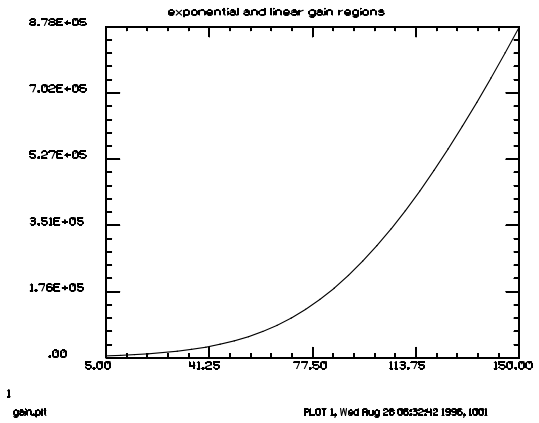


Fig. 9.1. Plot of energy growth as a function of distance. Initially the energy grows exponentially. In the saturated regime the energy grows linearly.

gain may be defined within a region that has a transverse function based on a supergaussian distribution. The transverse small signal gain is given a supergaussian spatial distribution of the form

$$g_0(x,y)\exp\left[-\left(\frac{x^2}{R_x^2}\right)^{N_x}-\left(\frac{y^2}{R_y^2}\right)^{N_y}\right], \tag{9.4}$$

where R_x and R_y are the supergaussian radii and N_x and N_y are the supergaussian exponents for the x- and y-directions. If R_x and R_y arcOe much larger than the region occupied by the beam, the gain is essentially unvarying.

Gain, with transverse variation, takes the form

$$I(x,y,z+\Delta z) \;=\; I(x,y,z)\exp\left[\frac{g_0\Delta z}{1+\dfrac{I(x,y,x)}{I_{sat}}}\right]. \tag{9.5}$$

9.2 Franz-Nodvik, CO_2 Gain

Franz-Nodvik kinetics for CO_2 pulsed lasers are called by the `gainco2` command.

$$I(x,y,z+\Delta z) \;=\; I_{sat}\ln\left\{1+\exp g_0(x,y)\exp\left[\left(\frac{I(x,y,z)}{I_{sat}}\right)-1\right]\right\}, \tag{9.6}$$

where Ln[ ] is the natural logarithm. Franz-Nodvik gain and Beer's Law gain are compared in Fig. 9.2.

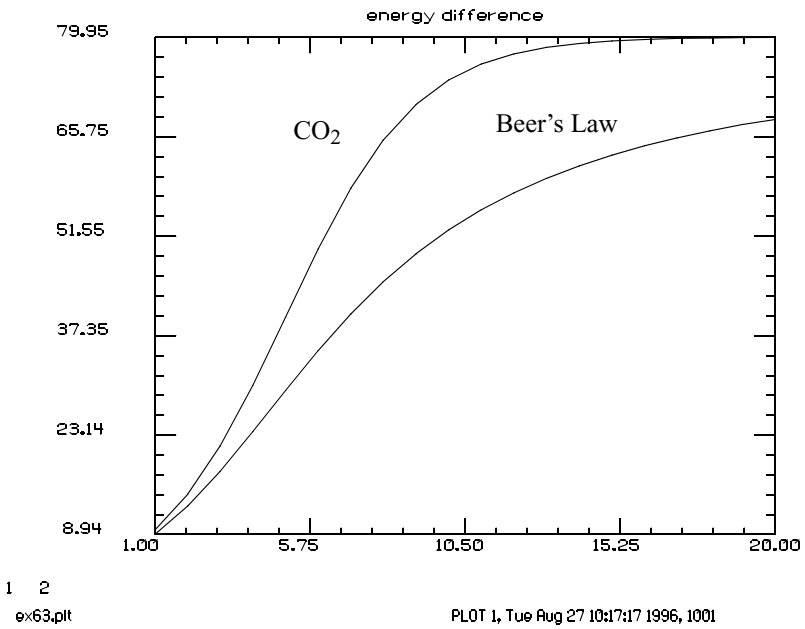


Fig. 9.2. Comparison of Beer's Law and pulsed CO_2 gain using Franz-Nodvik theory (from Example 63).

9.3 Rate Equation Model

A more detailed gain model based on the two-level atom and the rate equation approximation has been included. The state of some active medium may be characterized by the density of the medium and the population inversion. The complex amplitude arrays in GLAD may be used to store the populations density of the upper and lower level in the real and imaginary parts of the array. The array will constitute a gain sheet which represents an incremental axial segment of the gain medium. The gain sheet represents a volume of space of transverse dimensions $M\Delta x \times N\Delta y \times \Delta z$, where M is the number of x-elements in the array, N is the number of y-elements in the array, and Δz is the axial length of the array. Energy is pumped into the gain sheet by whatever pump method is used and energy decays from the gain sheet due to stimulated and spontaneous emission.

A four-level treatment of gain is one of the most commonly used models, as illustrated in Fig. 9.3. The rate equations are [4]

$$\Delta N_2 = \left[R_2 - \frac{N_2}{t_2} - (N_2 - N_1) W_i(\nu) \right] \Delta t, \tag{9.7}$$

$$\Delta N_1 = \left[R_1 - \frac{N_1}{t_{10}} + \frac{N_2}{t_{spont}} + (N_2 - N_1) W_i(\nu) \right] \Delta t, \tag{9.8}$$

where

ΔN_1 change in population of lower level, atoms/cm^3,

ΔN_2 change in population of upper level, atoms/cm^3,

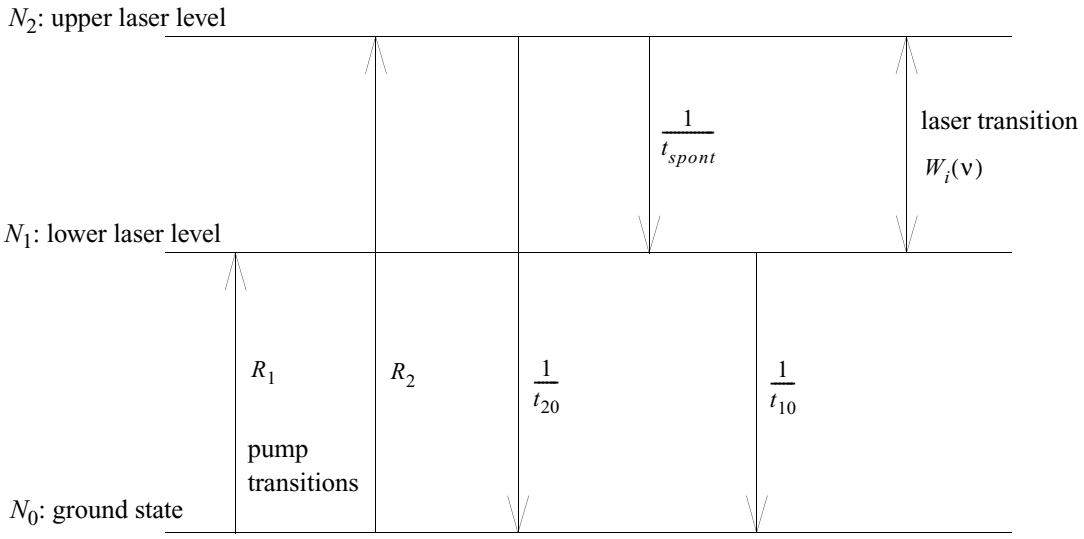


Fig. 9.3. Energy transitions for a four-level atom, R_1 and R_2 are the pump rates for Levels 1 and 2. t_{spont} is the lifetime for transition between Levels 1 and 2. t_{10} is the lifetime for decay from Level 1 to the ground state. t_{20} is the transition lifetime from Level 2 to the ground state.

R_2	pump rate for upper level, excitations/sec/cm^3,
R_1	pump rate for lower level, excitations/sec/cm^3,
t_{10}	decay time from lower level to ground, sec,
t_2	decay time from upper level to ground, sec, $1/t_2 = 1/t_{20} + 1/t_{spont}$,
t_{spont}	spontaneous decay lifetime, sec,
$W_i(\nu)$	transition probability density, probability/sec/cm^3,
Δt	elapsed time.

The transition probability density is,

$$W_i(\nu) = \frac{\lambda^2 g(\nu)}{8\pi n^2 h\nu t_{spont}} I, \tag{9.9}$$

where

λ	wavelength,
ν	normalized lineshape,
n	index of refraction,
h	Planck's constant,
ν	frequency of the radiation,
I	irradiance of the radiation.

The transition probability of Eq. (9.9) may be written in terms of the Einstein B-coefficient:

$$W_i(\nu) = B(\nu)\frac{I}{h\nu}, \text{ where } B(\nu) = \frac{\lambda^2 g(\nu)}{8\pi n^2 h\nu t_{spont}} I. \tag{9.10}$$

The small signal amplification takes the form

$$I(z) = I(0)e^{B\Delta N z}. \tag{9.11}$$

Under steady-state conditions, the irradiance of the optical field is constant and Eqs. (9.7) and (9.8) lead to the steady-state solution for the population inversion:

$$\Delta N^0 = R_2 t_2 - \left(R_1 + \frac{t_2}{t_{spont}}\right) R_2 t_{10}. \tag{9.12}$$

The small signal gain coefficient is

$$\gamma_0(\nu) = B(\nu)\Delta N^0. \tag{9.13}$$

The gain coefficient for homogeneous broadening and for arbitrary irradiance magnitude is,

$$\gamma(\nu) = \frac{\gamma_0(\nu)}{1 + \frac{I}{I_s}}, \tag{9.14}$$

where,

$$I_s = \frac{8\pi n^2 h\nu}{\left(\frac{t_2}{t_{spont}}\right)\lambda^2 g(\nu)} = \frac{h\nu}{B(\nu)t_2}. \tag{9.15}$$

In the case of strong saturation, Eq. (9.14) is well approximated by

$$\frac{dI}{dz} \approx \gamma_0(\nu) I_s = (B\Delta N^0)\left(\frac{h\nu}{B(\nu)t_2}\right) = \frac{\Delta N^0 h\nu}{t_2}. \tag{9.16}$$

Where the pumping rate into the upper level R_2 dominates the process, Eqs. (9.12) and (9.16) give the saturated gain coefficient as

$$\frac{dI}{dz} = R_2 h\nu, \; \gamma(\nu) = R_2 h\nu, \tag{9.17}$$

showing, in the case of saturated steady-state gain, a linear growth of irradiance with distance based on the pumping flux density.

9.3.1 Franz-Nodvik Solution

In GLAD we treat the laser as consisting of a discrete amplifier and efficiency loss due to outcoupling and other factors. The equation of optical amplification in the YAG rod is,

$$\frac{\partial I(z)}{\partial z} = B\Delta N(z)I(z) \tag{9.18}$$

The population inversion at each point is driven by the transition probality,

$$\frac{\partial \Delta N}{\partial z} = -2\Delta N W_i(\nu) = -2\Delta N \frac{BI}{h\nu} \tag{9.19}$$

In a coordinate system moving with the optical field, we make the change of variables $t = zn/c$

$$\frac{\partial \Delta N(z)}{\partial z} = \frac{-2n}{h\nu c} B\Delta N(z)I(z) \tag{9.20}$$

where $I(z)$ is the irradiance, $N(z)$ is the population inversion, and n is the index of the medium. The constant B is the cross section and has the value,

$$B = \frac{\lambda^2}{8\pi t_{\text{spont}}} g(\nu) \tag{9.21}$$

where λ is the wavelength in the medium (not the vacuum wavelength) and ν is area-normalized spectral response function,

$$g(\nu) = \frac{\Delta\nu}{2\pi\left[(\nu-\nu_0)^2 + \left(\frac{\Delta\nu}{2}\right)^2\right]} \tag{9.22}$$

Note that the ν has peak value $2/\pi$ because of the area normalization.

In the general case of a square pulse traveling through an amplifying medium, we have the Frantz-Nodvic solution,

$$I(\Delta) = I_s \text{Ln}\left\{1 + \exp\left(B\int_0^{\Delta z} \Delta N(z)dz\right)[\exp(I(0)/I_s) - 1]\right\} \tag{9.23}$$

where $I_s = h\nu/B$, as illustrated in Fig. 80.2(a). this gives $\Delta N(z)$ as a continuously varying function. In numerical calculation, we must treat $\Delta N(z)$ as a set of discrete gain sheets and we can not differentiate the number density axially within a gain sheet, as shown in Fig. 80.2(b). Withing a gain sheet we must assume, that all of the population inversion along an axial line intersects with all of the light for a pulse of Δt. For a laser where the population inversion in the gain sheets is much more uniform because optical beams are entering from both sides, as illustrated in Fig. 80.2(c).

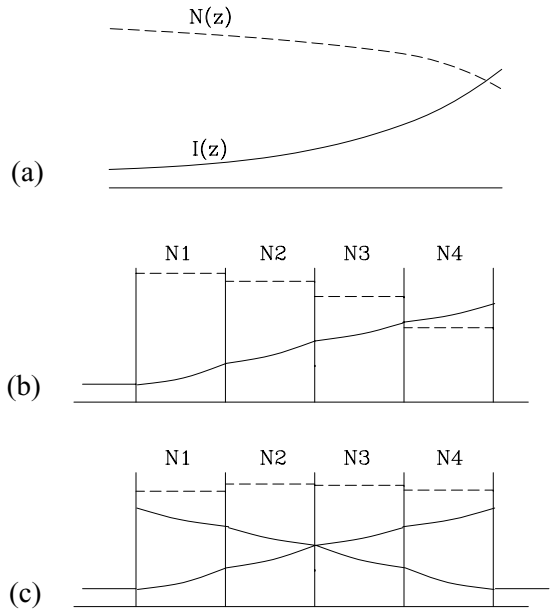


Fig. 9.4. (a). Continuous solution for and for an optical beam passing through an amplifier. (b) shows a gain sheet solution where all of the population inviersion in each gain sheet is assumed to interact as a unit. (c) shows more uniform population inversion in the gain sheet when the amplifier is used in double pass in the typical laser.

Equations (9.18) and (9.19) may be solved by the standard numerical methods for solving coupled differential equations. Unfortunately this approach is too slow since it may be necessary to solve the rate equations millions of times in the course of an analysis. An approximate solution is to take advantage of photon conservation to simplify Eqs. (9.18) and (9.19). We then have the approximate solution,

$$I(z) = I(0)e^{B\Delta\Delta N(0)\Delta z} \tag{9.24}$$

$$\Delta N(z) = \Delta N(0) - \frac{2n}{h\nu c}[I(z) - I(0)] \tag{9.25}$$

Equations (9.24) and (9.25) may be executed very quickly — requiring only the evaluation of a single exponential. These equations work very well for many rate equation calculations such as laser startup and steady-state amplification. However, the gain in a Q-switch device is much higher than can generally be achieved during laser startup. Under conditions of extremely high gain Eq. (9.24) over-predicts the amplification because the initial value of ΔN is used in Eq. (9.24). Since the amplification is somewhat too

high, the depletion of the population inversion is somewhat too rapid. With very great amplification, the error may cause Eq. (14) to predict a negative population inversion, which can not arise from stimulated emission. This is a fairly intractable problem for extreme gain and efforts to provide numerical "patches" have been less than satisfactory. Through the standard method of numerical solution of coupled differential equations we have a robust method that is slow. The method of Eqs. (13) and (14) is fast but not robust.

A method that is both fast and robust is possible by reevaluating the problem. Eqs (9.18) and (9.20) are appropriate for a small temporal sample of a beam traveling through an optical amplifier. In a resonator, the gain medium interacts with the entire optical field in the device. We use a single computer array (or at most two, for the two polarizations) to represent the entire optical field in the resonator. We can not, therefore, distinguish temporal events which occur on a scale less than the round-trip time of the resonator. We could, in principle, use multiple temporal samples to resolve time events shorter than the round-trip time, but this is not necessary for the Q-switch study.

If we consider the optical field to be of intensity I and of duration, Δt, the round-trip time: then the optical field contains a well-defined photon flux. The potential photon flux increase due to the population inversion is

$$\frac{1}{2}\Delta N(0)L \tag{9.26}$$

where L is the length of the gain region, as illustrated in Fig. 9.5. The net energy of an incident square pulse of irradiance $I(z)$ and temporal length Δt, giving the energy density as $I(0)\Delta t$. The energy density in a gain sheet representing a length of L is $\Delta N(0)h\nu L/2$. The sum of these two energy densities is a constant by conservation of energy.

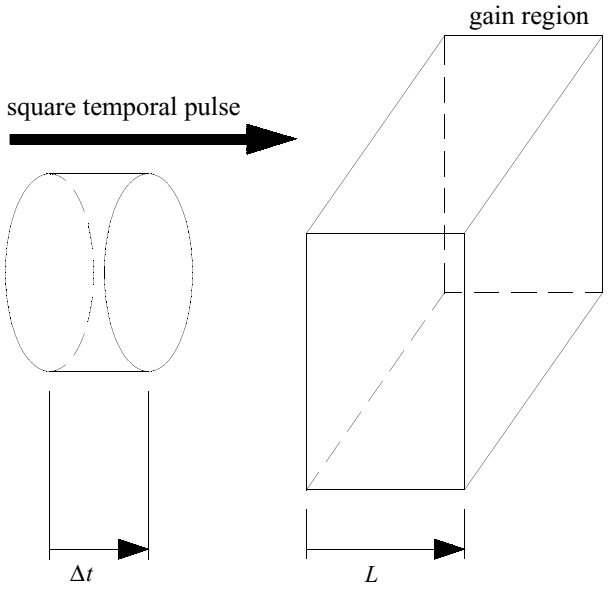


Fig. 9.5. A time slice of the pulse of length interacts with a gain region of length.

$$\text{total energy density} = I(0) + \frac{1}{2}\Delta N(0) h\nu L \tag{9.27}$$

By dividing by Δt, we can calculate the maximum possible irradiance if all the population inversion were transformed into light,

$$I_{\max} = I(0) + \frac{\Delta N(0) h\nu L}{2\Delta t} \tag{9.28}$$

and by dividing by $h\nu L/2$, we have the maximum population inversion if all the light were subsumed by stimulated absorption,

$$\Delta N_{\text{total}} = \Delta N(0) + \frac{2I(0)\Delta t}{h\nu L} \tag{9.29}$$

We can use ΔN_{total} to calculate $N(z)$,

$$\Delta N(z) = \Delta N_{\text{total}} - \frac{2I(0)\Delta t}{h\nu L} \tag{9.30}$$

Equation (9.18) now takes the form,

$$\frac{\partial I(z)}{\partial z} = B\left[\Delta N_{\text{total}} - \frac{2I(z)\Delta t}{h\nu L}\right] I(z) \tag{9.31}$$

Equation (20) has the exact solution,

$$I(L) = \frac{I_{\max} I(0)}{I(0) + (I_{\max} - I(0)) e^{-B\Delta N_{\text{total}} L}} \tag{9.32}$$

At low saturation, Eq. (9.32) approaches the expected simple exponential gain. At high saturation, Eq. (9.32) approaches $I_{\max}$. Eq. (9.32) works well for both high and low energy amplifiers and the rate equation algorithms have been modified to use this expression.

9.3.2 Off-line Effects

The gain and off-line index of refraction effects may be represented by a complex index of refraction using χ'_m and χ''_m such that

$$n \to n\left(1 + \frac{\chi'_m}{2n^2} + \frac{\chi''_m}{2n^2}\right), \tag{9.33}$$

$$\chi''_m = (N_1 - N_2)\frac{\lambda^3}{16\pi^3 t_{spont} n} g(\nu), \quad \chi'_m = \frac{2(\nu_{off} + m\Delta\nu_c)}{\Delta\nu}\chi''_m, \tag{9.34}$$

$$g(\nu_m) = \frac{\Delta\nu}{2\pi\left[(\nu_{off} + m\Delta\nu_c)^2 + \left(\frac{\Delta\nu}{2}\right)^2\right]}, \tag{9.35}$$

where $\nu_m - \nu_{cen} = \nu_{off} + m\Delta\nu_c$, and m is the mode number. The optical field, under steady state conditions varies as

$$a_m(x, y, \Delta t) = a_m(x, y, 0) e^{(jk\chi'_m + k\chi''_m)\frac{L}{2n^2}}, \tag{9.36}$$

where $\chi_m = \chi'_m - j\chi''_m$ is the electric susceptibility and n is the index of refraction.

9.3.3 Spontaneous Emission

Spontaneous emission arises from the decay of the population inversion. This spontaneous emission is a noise source for many laser processes. The noise power injected into each mode in a distance Δz is

$$\Delta I_{\text{noise}} = \frac{(N_2 - N_1)h\Delta z}{2t_{spont}} \frac{\lambda^2}{4\pi\Delta x\Delta y}, \tag{9.37}$$

where the solid angle subtended by the computer array is $\Delta\Omega = \lambda^2 / 4\pi\Delta x\Delta y$ when using sampling intervals of Δx and Δy. This noise is introduced as a delta-correlated, normally distributed random phasor.

9.4 Three-Level Gain, Ruby Laser

The ruby laser, invented by Maiman, was the first working laser[6]. A study by Kuo, et. al., describes numerical methods of the Q-switched ruby laser[7]. The gain of the ruby laser depends only on the populations of two states[8]. The upper state is split into two levels, $2A$ and E. Their separation is 29 cm^{-1} (870 GHz, 1.4 nm). The higher $2A$ state to ground emits at 692.2 nm and the E state to ground emits at 694.3 nm. The E to ground transition has the higher gain because its population is 15% higher in thermal equilibrium at $300°\text{K}$. They are connected so closely by thermal equilibration that the energy in E is effectively the total energy in the two split states. The gain depends on the population in the upper state E, the lower state N_1, and each state's degeneracy $g_d(\ )$. The degeneracies are: $g_d(n_1) = 4$, $g_d(2A) = 2$, and $g_d(E) = 2$.

The net population of the split upper level is,

$$N_2 = N_2(2A) + N_2(E). \tag{9.38}$$

The total population of the upper and lower states is constant on a pointwise basis:

$$N_{tot} = N_1 + N_2 = N_1 + N_2(2A) + N_2(E) = \text{constant for each x, y point,} \tag{9.39}$$

N_{tot} depends only on the doping density in the crystal. The ratio of the population densities between the split levels depends on the Boltzman distribution K

$$K = \frac{N_2(2A)}{N_2(E)} = \exp\left(-\frac{\Delta E}{kT}\right) \approx 0.87 \ @ \ 300°K. \tag{9.40}$$

The gain is

$$\gamma(\nu) = \sigma_{21}\left[N_2(E) - \frac{g_d(\nu)}{g_d(N_1)}N_1\right], \tag{9.41}$$

where $\sigma_{21}(\nu)$ is the cross section between levels E and N_1.

For ruby we have the relationship between $\sigma_{21}(\nu)$ and $\sigma_{12}(\nu)$,

$$\sigma_{21}(\nu) = \frac{g_d(N_1)}{g_d(E)}\sigma_{12}(\nu) = 2\sigma_{12}(\nu). \tag{9.42}$$

$\sigma_{21}(\nu)$ is

$$\sigma_{21}(\nu) = \frac{\lambda^2}{8\pi n^2 t_{21}} g(\nu), \tag{9.43}$$

$$g(\nu) = \frac{\Delta(T)}{2\pi} \frac{1}{(\nu - \nu_0)^2 + \left(\frac{\Delta\nu(T)}{2}\right)^2}. \tag{9.44}$$

Typical values are $t_{21} = 0.00511 - 1.057\times10^{-5}(T - 293°K)$ sec and $\Delta\nu(300°K) = 2\times10^{11}$ Hz. From Eqs. (9.23) through (9.26) we have

$$\gamma(\nu) = \sigma_{21}(\nu)\left[\frac{3+K}{2(1+K)}N_2 - \frac{N_{tot}}{2}\right]. \tag{9.45}$$

The rate equations for interaction of population inversion and optical field ϕ, as photon density, are

$$\frac{\partial N_2}{\partial t} = \frac{RN_1}{h\nu} - \frac{N_2}{t_{21}} - \gamma(\nu)\phi, \tag{9.46}$$

$$\frac{\partial \phi}{\partial t} = \phi_{\text{noise}} - \frac{\phi}{t_c}, \tag{9.47}$$

$$\frac{\partial \phi}{\partial z} = \gamma(\nu)\phi, \tag{9.48}$$

where t_c represents the cavity loss and ϕ_{noise} represents the contribution from spontaneous emission. The pumping rate R is the product of the pump irradiance and the pumping cross section. The cavity loss may be incorporated separately by representing the aperture clipping and other losses, so it need not be solved explicitly in the pair of differential equations. The spontaneous emission and pumping rate may be separated out leaving the optical field and population inversion interaction

$$\frac{\partial N_2}{\partial t} = -\gamma(\nu)\varphi = -\varphi[\alpha(\nu)N_2 - \beta(\nu)N_{tot}] = -\alpha(\nu)\varphi\left[N_2 - \frac{\beta(\nu)}{\alpha(\nu)}N_{tot}\right], \tag{9.49}$$

where $\alpha(\nu) = \sigma_{21}(\nu)\frac{3+K}{2(1+K)}$, $\beta(\nu) = \frac{\sigma_{21}(\nu)}{2}$,

$$\frac{\partial \varphi}{\partial z} = \gamma(\nu)\phi = \varphi[\alpha(\nu)N_2 - \beta(\nu)N_{tot}] = \alpha(\nu)\phi\left[N_2 - \frac{1+K}{3+K}N_{tot}\right]. \tag{9.50}$$

We may make the substitutions $N' = N_2 - \frac{\beta}{\alpha}N_{tot}$ and $\phi = \frac{I}{h\nu}$,

$$\frac{\partial}{\partial t}N' = -\alpha(\nu)\frac{I}{h\nu}N', \tag{9.51}$$

$$\frac{\partial I}{\partial z} = \alpha(\nu)N'I. \tag{9.52}$$

Equations (9.51) and (9.52) may be solved by the Franz-Nodvik method for the important special case of a square temporal pulse traveling through an amplifying medium, as illustrated in Fig. 9.5.If we consider the optical field to be of intensity I and of duration, Δt, the round-trip time; then the optical field contains a well-defined photon flux. The potential photon flux increase due to the population inversion is

$$N'(0)L, \tag{9.53}$$

where L is the length of the gain region. The net energy of an incident square pulse of irradiance $I(z)$ and temporal length Δt, giving the energy density as $I(0)\Delta t$. The energy density in a gain sheet representing a length of L is $N'(0)h\nu L$. The sum of these two energy densities is a constant by conservation of energy.

$$\text{total energy density} = I(0)\Delta t + N'(0)h\nu L. \tag{9.54}$$

By dividing by Δt, we can calculate the maximum possible irradiance if all the population inversion were transformed into light,

$$I_{\max} = I(0) + \frac{N'h\nu L}{\Delta t}, \tag{9.55}$$

and by dividing by $h\nu L$, we have the maximum population inversion if all the light were subsumed by stimulated absorption,

$$N'_{\max} = N'(0) + \frac{I(0)\Delta t}{h\nu L}. \tag{9.56}$$

We can use $N'_{\max}$ to calculate $N'(z)$,

$$N'(z) = N'_{\max} - \frac{I(z)\Delta t}{h\nu L}. \tag{9.57}$$

Equation (9.52) now takes the form,

$$\frac{\partial}{\partial z} I(z) = \alpha(\nu)\left(N'_{\max} - \frac{I(z)\Delta t}{h\nu L}\right) I(z). \tag{9.58}$$

Equation (9.58) has the exact solution,

$$I(L) = \frac{I_{\max} I(0)}{I(0) + \left(I_{\max} - I(0) e^{-\alpha(\nu) N'_{\max} L}\right)}, \tag{9.59}$$

$$\Delta I = I(L) - I(0), \tag{9.60}$$

$$\Delta N_2 = \Delta N'_2 = -\frac{\Delta I \Delta t}{h\nu l}. \tag{9.61}$$

At low saturation, Eq. (9.59) approaches the expected simple exponential gain. At high saturation, Eq. (9.59) approaches $I_{\max}$. Eq. (9.59) works well for both high and low energy amplifiers and the rate equation algorithms in GLAD take advantage of this simplification.

9.4.1 Sample Calculation: Three-level Gain, Ruby Laser

Assume the following values:

linewidth $\Delta\nu(293°\text{K}) = 3.3\times10^{11}\,\text{sec}^{-1}$

decay time $t_{21}(293°\text{K}) = 0.00511\,\text{sec}$

T = 293 ° K

$\lambda(293°\text{K}) = 0.694325\mu$, $\nu_0(293°\text{K}) = 4.3178\times10^{14}$

$$\sigma_{21}(\nu_0) = \frac{\lambda^2}{8\pi n^2 t_{21}} g(\nu_0) = \frac{\lambda^2}{8\pi n^2 t_{21}\Delta\nu} = 2.31145\times10^{-20}\text{cm}^2$$

$$N_{tot} = 1.58\times10^{19}\text{cm}^{-3}$$

$$\frac{\Delta E}{k} \approx 41.74\,.\; K = \exp\left(-\frac{\Delta E}{kT}\right) = 0.8671062 \text{ for T} = 293^\circ\text{K}.$$

$$\alpha = \sigma_{21}\frac{3+K}{2(1+K)} = 2.39371\times10^{-20}\text{cm}^2,\ \frac{\beta}{\alpha} = \frac{1+K}{3+K} \approx 0.48\text{, and } N' = N_2 - 0.4828174N_{tot}$$

For small signal gain, $N_1 \approx 0$ and $N_2 \approx N_{tot}$ which implies the small signal gain is

$$\gamma_0 = \sigma_{21}\left[\frac{3+K}{2(1+K)}N_2 - \frac{N_{tot}}{2}\right] \approx 0.5171826\sigma_{21}N_{tot} \approx 0.1956017\text{cm}^{-1}$$

$$N'(0) = 0.4828174N_{tot} = 0.8171486\times10^{19}\text{cm}^{-3}.$$

The contribution of $N'(0)$ to the maximum intensity is $N'(0)h\nu L/(\Delta t)$. Given $h = 6.626\times10^{-34}$ and $c = 2.99792458\times10^{10}$cm/sec, $h\nu = 2.860945\times10^{-19}$. Given a pulse width of $\Delta t = 0.1$nsec and length of $L = 1$cm, the contribution of $N(0)$ to the maximum intensity is $I_{\max} = N'(0)h\nu L/\Delta t = 2.33782\times10^{10}\text{w/cm}^2$. $I_{\max}$ represents the saturation value. For a starting irradiance of $I(0) = 10^{-4}$,

$$I(L = 1\text{cm}) = \left.\frac{I_{\max}I(0)}{I(0) + (I_{\max} - I(0))e^{-\alpha(\nu)N'_{\max}L}}\right|_{L\,=\,1\text{cm}} = 1.216042\times10^{-4}\text{w/cm}^2,$$

$$\Delta I = I(L) + I(0) = 0.216042\times10^{-4}\text{w/cm}^2,$$

$$\Delta N_2 = \Delta N_2' = \frac{\Delta I\Delta t}{h\nu L} = 7551.438\text{photon/cm}^3.$$

9.5 Three-Level Gain, Single Upper State

In the case of a three-level system with a single upper (non-split) state, the lasing is between the upper state and ground state. The total population of the upper and lower states is constant on a pointwise basis:

$$N_{tot} = N_1 + N_2 = \text{constant for each } x, y \text{ point.} \tag{9.62}$$

The gain is

$$\gamma(\nu) = \sigma_{21}(\nu)[N_2 - N_1] = \sigma_{21}(\nu)[2N_2 - N_{tot}], \tag{9.63}$$

where $\sigma_{21}(\nu)$ is the cross section between levels N_2 and N_1. $\sigma_{21}(\nu)$ is

$$\sigma_{21}(\nu) = \frac{\lambda^2}{8\pi n^2 t_{21}} g(\nu), \tag{9.64}$$

$$g(\nu) = \frac{\Delta\nu(T)}{2\pi} \frac{1}{(\nu - \nu_0)^2 + \left(\frac{\Delta\nu(T)}{2}\right)^2}. \tag{9.65}$$

A typical value for an excimer using KrF is $t_{21} = 8.1\times10^{-9}\text{sec}$. The rate equations for interaction of population inversion and optical field ϕ, as photon density, are

$$\frac{\partial N_2}{\partial t} = \frac{RN_1}{h\nu} - \frac{N_2}{t_{21}} - \gamma(\nu)\phi, \tag{9.66}$$

$$\frac{\partial \phi}{\partial t} = \varphi_{\text{noise}} - \frac{\phi}{t_c}, \tag{9.67}$$

$$\frac{\partial \phi}{\partial z} = \gamma(\nu)\phi, \tag{9.68}$$

where t_c represents the cavity loss and ϕ_{noise} represents the contribution from spontaneous emission. The pumping rate R is the product of the pump irradiance and the pumping cross section. The cavity loss may be incorporated separately by representing the aperture clipping and other losses, so it need not be solved explicitly in the pair of differential equations. The spontaneous emission and pumping rate may be separated out leaving the optical field and population inversion interaction

$$\frac{\partial N_2}{\partial t} = -\gamma(\nu)\phi = -\phi\sigma_{21}(\nu)[2N_2 - N_{tot}]. \tag{9.69}$$

We may make the substitutions $N' = 2N_2 - N_{tot}$ and $\phi = I/h\nu$,

$$\frac{\partial}{\partial t}N' = -\sigma_{21}(\nu)\frac{I}{h\nu}N', \tag{9.70}$$

$$\frac{\partial I}{\partial z} = \sigma_{21}(\nu)N'I. \tag{9.71}$$

Equations (9.70) and (9.71) may be solved by the Franz-Nodvik method for the important special case of a square temporal pulse traveling through an amplifying medium.

If we consider the optical field to be of intensity I and of duration Δt, the round-trip time, then the optical field contains a well-defined photon flux. The potential photon flux increase due to the population inversion is

$$N'(0)L\,, \tag{9.72}$$

where L is the length of the gain region. The net energy of an incident square pulse of irradiance $I(z)$ and temporal length Δt, giving the energy density as $I(0)\Delta t$. The energy density in a gain sheet representing a length of L is $N'(0)h\nu L$. The sum of these two energy densities is a constant by conservation of energy.

$$\text{total energy density} = I(0)\Delta t + N'(0)h\nu L\,. \tag{9.73}$$

By dividing by Δt, we can calculate the maximum possible irradiance if all the population inversion were transformed into light,

$$I_{\max} = I(0) + \frac{N'h\nu L}{\Delta t} \tag{9.74}$$

and by dividing by $h\nu L$, we have the maximum population inversion if all the light were subsumed by stimulated absorption,

$$N'_{\max} = N'(0) + \frac{I(0)\Delta t}{h\nu L}\,. \tag{9.75}$$

We can use $N'_{\max}$ to calculate $N'(z)$,

$$N'(z) = N'_{\max} - \frac{I(z)\Delta t}{h\nu L}\,. \tag{9.76}$$

Equation (9.71) now takes the form,

$$\frac{\partial}{\partial z}I(z) = \alpha(\nu)\left(N'_{\max} - \frac{I(z)\Delta t}{h\nu L}\right)I(z)\,. \tag{9.77}$$

Equation (9.77) has the exact solution,

$$I(L) = \frac{I_{\max}I(0)}{I(0) + [I_{max} - I(0)]e^{-\alpha\nu N'_{\max}L}}\,, \tag{9.78}$$

$$\Delta I = I(L) - I(0)\,, \tag{9.79}$$

$$\Delta N_2 = \Delta N'_2 = -\frac{\Delta I \Delta t}{h\nu l}. \tag{9.80}$$

At low saturation, Eq. (9.78) approaches the expected simple exponential gain. At high saturation, Eq. (9.78) approaches $I_{\max}$. Eq. (9.78) works well for both high and low energy amplifiers and the rate equation algorithms in GLAD take advantage of this simplification.

Medium pumping is found from the rate equations:

$$\frac{dN_2}{dt} = RN_1 - \frac{N_2}{t_2}, \tag{9.81}$$

$$\frac{dN_2}{dt} = R(N_{tot} - N_2) - \frac{N_2}{t_2}, \tag{9.82}$$

$$\frac{dN_2}{dt} = RN_{tot} - \left(R + \frac{1}{t_2}\right)N_2, \tag{9.83}$$

$$N_2(t) = \frac{RN_{tot}}{R + \frac{1}{t_2}} + \left(N_2(0) - \frac{RN_{tot}}{R + \frac{1}{t_2}}\right)\exp\left[-\left(R + \frac{1}{t_2}\right)t\right]. \tag{9.84}$$

The steady-state upper state population is,

$$N_{ss} = \frac{Rt_2 N_{tot}}{1 + Rt_2}, \tag{9.85}$$

$$N_2(t) = N_{ss} + (N_2(0) - N_{ss})\exp\left[-\left(\frac{Rt_2 + 1}{t_2}\right)t\right], \tag{9.86}$$

$$N_2(t) = (N_{ss} - N_2(0))\left\{1 - \exp\left[-\left(\frac{Rt_2 + 1}{t_2}\right)t\right]\right\}. \tag{9.87}$$

We need to have $N_{ss} > N_{tot}/2$ to have a population inversion, so it is necessary to have $Rt_2 > 1$. The critical pumping rate is $Rt_2 = 1$.

9.5.1 Multiple Mode Interactions

Given several modes identified by the k-index and having I_k, frequencies ν_k, and associated cross sections σ_k, all modes complete for the same population inversion. If all optical modes are treated as

accessing the population inversion independently, then the gain will be calculated to be too high. This will be evident because the sum of real and potential irradiance will be increased by the calculation — a violation of conservation of energy.

Total real and potential irradiance

$$I_{tot}(0) = \frac{N(0)h\nu L}{\Delta t} + \sum_{k=1}^{N} I_k(0), \tag{9.88}$$

$$I_{k_{\max}} = I_k(0) + \frac{\alpha N'(0)h\nu L}{\Delta t}, \tag{9.89}$$

$$N'_{k_{\max}} = \alpha N'(0) + \frac{I_k(0)\Delta t}{h\nu L}, \tag{9.90}$$

$$I_k(L) = \frac{I_{k\max} I(0)}{I_k(0) + [I_{k\max} - I(0)]e^{-\sigma_{21}\nu N'_{k\max} L}}. \tag{9.91}$$

We compute the real and potential irradiance at the end

$$I_{tot}(L) = \frac{N(L)h\nu L}{\Delta t} + \sum_{k=1}^{N} I_k(L). \tag{9.92}$$

This value will exceed, at least by a small amount, because the population inversion will effectively be allocated to all modes. We may estimate the value of α from the expression:

$$\alpha \approx 1 - \frac{\Delta t\left(\sum_{k=1}^{N} I_k(L) - \sum_{k=1}^{N} I_k(0)\right)}{N(0)h\nu L}. \tag{9.93}$$

One estimates the correction factor α and recalculates the values $I_k(L)$ and a new value for total irradiance. By iteration of the process, one finds the value of α that conserves energy.

9.5.2 Sample Calculation: Three-level Gain, Single Upper Level

Consider a KrF excimer laser. Assume the following values:

$$\nu_0 = \frac{c}{\lambda} = 1.207\times10^{15}$$

$h\nu = 7.998\times10^{-10}\text{cm}^{-3}$

linewidth is 50 picometers, $\Delta\nu = 50\times10^{-10}\frac{\nu}{\lambda} \approx 2.43\times10^{11}\text{Hz}$

spontaneous decay time $t_{spont} = 8.1\times10^{-9}\text{sec}$

index of refraction $n = 1$

$\lambda = 0.24835\mu$, $\nu = 1.208\times10^{15}$,

$$g(\nu) = \frac{\Delta\nu}{2\pi}\frac{1}{(\nu-\nu_0)^2+\left(\frac{\Delta\nu}{2}\right)^2}$$

for $\nu = \nu_0$, $g(\nu) = \frac{2}{\pi\Delta\nu} = 2.62\times10^{-12}$

$$\sigma_{21}(\nu_0) = \frac{\lambda^2}{8\pi n^2 t_{spont}}g(\nu_0) = \frac{\lambda^2}{8\pi n^2 t_{spont}}\frac{2}{\pi\Delta\nu} = 7.937\times10^{-15}\text{cm}^2$$

Select maximum energy density to be $E_{tot} = 2\times10^{-5}\text{Jcm}^{-3}$ yielding a maximum population of $N_{tot} = \frac{J_{tot}}{h\nu} = 2.5\times10^{13}\text{cm}^{-3}$. For complete inversion we would be able to extract a maximum of 0.001 J cm-2. The maximum possible small signal gain is $\gamma_{0_{\max}} = \sigma_{21}N_{tot} = 0.1984\text{cm}^{-1}$.

The spontaneous decay time is $t_{spont} = 8.1\times10^{-9}$ sec. Assuming a non-spontaneous decay rate of $t_{21} = 8.1\times10^{-8}$ sec, the net decay rate is $t_2 = \frac{1}{\frac{1}{t_{spont}}+\frac{1}{t_{21}}} = 7.364\times10^{-9}$ sec.

Select the pumping rate to be $R = \frac{2}{t_2}$. $N_{ss} = \frac{2}{3N_{tot}} = 1.667\times10^{13}$ cm-3. For a pumping time of $t = 4\times10^{-9}$ sec., the population is $\Delta N_2(t) = (N_{ss} - N_2(0))\left[1 - \exp\left(-\frac{Rt_2+1}{t_2}t\right)\right]$. For a medium initially at rest $N_2(0) = 0$,

$$\Delta N_2(t) = (1.667\times10^{13} - 0)\left[1 - \exp\left(-\frac{3}{7.364\times10^{-9}}4\times10^{-9}\right)\right]$$

$$= \left(0.804N_{ss} = \frac{2}{3}(0.804)N_{tot} = 0.536N_{tot}\right).$$

The population inversion after this short interval of pumping is, therefore, $\Delta N = 0.072N_{tot}$. The small signal gain is

$$\gamma_0 = \sigma_{21}\Delta N \approx 0.7142\text{cm}^{-1} = (0.072)(0.1984)\text{cm}^{-1} = 1.428\times10^{-2}\text{cm}^{-1}.$$

The small signal gain length is $\gamma_0 L = 0.7142$ cm-1. The noise power is

$\langle \Delta I_{\text{noise}} \rangle = \frac{(N_2 - N_1)h\nu\Delta z}{2t_{spont}} \frac{\lambda^2}{4\pi\Delta x\Delta y}\left(\frac{1}{7}\right)$ where the factor of 1/7 includes the line narrowing typically associated with nonlinear gain. Including the numeric values we have,

$$6.626\times10^{34} \cdot 1.207\times10^{15} \frac{50}{2 \cdot 8.1\times10^{-9}}(0.303)\frac{1}{7} = 1.73\times10^{-18}(N_2 - N_1) = 3.11\times10^{-6}\text{w cm}^{-2}.$$

Time per pass = 4×10^{-9} sec. To achieve approximately 0.01 J per pulse out of a gain region of 50 cm, we need an energy density of about $J_{tot} = (0.001/50)\text{J} \cdot \text{cm}^{-1}$, so that

$N_{tot} = 0.001/50/h\nu = 2.597\times10^{13}\text{cm}^{-3}$.

To pump at least twice the inverse of the net level decay rate, assume $R = 2/t_{spont} = 2.469\times10^{8}\text{cm}^{-1}$.

$N'(0) = 0.4828174N_{tot} = 0.8171486\times10^{19}\text{cm}^{-3}$. The contribution of $N'(0)$ to the maximum intensity is $I_{max} = \frac{N'(0)h\nu L}{\Delta t} = 2.33782\times10^{10}\text{w/cm}^2$. I_{max} represents the saturation value. For a starting irradiance of $I(0) = 10^{-4}$,

$$I(L = 1\text{cm}) = \left.\frac{I_{\max}I(0)}{I(0) + [I_{max} - I(0)]e^{-\alpha\nu N'_{\max}L}}\right|_{L = 1\text{cm}} = 1.216042\times10^{-4}\text{w/cm}^2$$

$\Delta I = I(L) - I(0) = 0.216042\times10^{-4}\text{w/cm}^2$.

9.6 Raman Modeling

In its most general form, Raman interactions can be highly complex. It may be necessary to treat transient effects, multiple longitudinal modes, multiple Stokes orders, four-wave mixing effects, and spontaneous emission.

The original objective was to treat an excimer/Raman system characterized by a broad spectrum of longitudinal modes. The spectral characteristics of the pump and Stokes beam may be different and the gain of the Raman process depends on the spectral correlation. The treatment is facilitated by the reasonable assumption that the transverse characteristics of all modes is identical. The general multiple mode description, including nonresonant terms, is

$$\frac{\partial P_j}{\partial z} = -\frac{g}{2}\sum_n\sum_k \frac{P_k S^*_{k-n} S^*_{j-n}}{1+\frac{i2n\Delta}{\Gamma}} e^{i(j-k)\Delta\nu z}, \frac{\partial S_j}{\partial z} = \frac{g}{2}\sum_n\sum_k \frac{S_k P^*_{k-n} P^*_{j-n}}{1+\frac{i2n\Delta}{\Gamma}} e^{i(k-j)\Delta\nu z}, \tag{9.94}$$

where P_j and S_j are the photon flux complex amplitudes of the jth pump and Stokes modes (offset in frequency by the Raman shift), Γ is the homogeneously broadened line width, g is the small signal gain coefficient and Δ is the mode spacing. The indices n and k designate the nth nonresonant term and the kth mode respectively. The variable ν is a measure of the group velocity dispersion,

$$\nu = \frac{1}{\nu_p} - \frac{1}{\nu_s}, \tag{9.95}$$

where ν_p and ν_s are the group velocities of the pump and Stokes waves.

The photon flux complex amplitude is related to the irradiance complex amplitude by $\sqrt{f}$, where f is the optical frequency. For example the photon flux complex amplitude of the pump beam is

$$P = \frac{E_p}{\sqrt{hf}}, \tag{9.96}$$

where h is Plank's constant E_p is the complex amplitude used in GLAD.

In this notation the line width of a single mode is assumed to be zero and the mode spacings of pump and Stokes beams are identical. When finite spectral width is to be included, the summations are replaced with integrations over frequency.

If $\Delta \gg \Gamma$, there is substantial gain only when $n = 0$. This is the case of resonant interaction. Even in the case where we can not assume $\Delta \gg \Gamma$, nonresonant terms may be negligible (except for print through effects) if there are a sufficient number of modes because the nonresonant terms add incoherently to the resonant term. For broadband excimer/Raman interactions typical of the collinear Raman geometry, the nonresonant terms may be neglected. In that case, we may simplify Eq. (9.94) by setting $n = 0$. The simplified equations are

$$\frac{\partial P_j}{\partial z} = -\frac{g}{2}\sum_k S_j (P_k S^*_k), \frac{\partial S_j}{\partial z} = \frac{g}{2}\sum_k P_j (S_k P^*_k). \tag{9.97}$$

In more compact notation, the inner product may be written,

$$\frac{\partial P_j}{\partial z} = -\frac{g}{2} S_j (\vec{\tilde{S}}^* \cdot \vec{\tilde{P}}), \frac{\partial S_j}{\partial z} = \frac{g}{2} P_j (\vec{\tilde{P}}^* \cdot \vec{\tilde{S}}), \tag{9.98}$$

where

$\vec{\tilde{P}}$ = vector of all pump longitudinal mode complex amplitudes,

$\vec{\tilde{S}}$ = vector of all 1st Stokes longitudinal mode complex amplitudes,

$\vec{\tilde{S}}^* \cdot \vec{\tilde{P}} = \sum_k S_k^* P_k$.

If $\vec{\tilde{S}}$ and $\vec{\tilde{P}}$ consist of a single mode, Eq. (9.98) simplifies, in air, to

$$\frac{\partial P}{\partial z} = -\frac{g}{2} |S|^2 P, \frac{\partial S}{\partial z} = \frac{g}{2} |P|^2 S. \tag{9.99}$$

9.6.1 Photon-Flux Raman Model

A simple model of Raman scattering uses only the intensities of the seed and pump beam. This model neglects four-wave mixing, second-Stokes, and anti-Stokes effects. This model is commonly referred to as the photon-flux model.

From Eq. (9.99), an expression in terms of the photon flux parameters may be derived,

$$\frac{\partial |P|^2}{\partial z} = -g |S|^2 |P|^2, \frac{\partial |S|^2}{\partial z} = g |P|^2 |S|^2. \tag{9.100}$$

Using Eq. (9.96), we have the relation

$$\frac{I_s}{I_p} = \frac{\lambda_s}{\lambda_p} \frac{|S|^2}{|P|^2}, \tag{9.101}$$

where I_s and I_p are the seed and pump intensities and λ_s and λ_p are the seed and pump wavelengths. Using Eqs. (9.100) and (9.101) an expression for the intensity after propagation by a distance Δz, an analytical solution is possible[3], as shown below,

$$I_s(x,y,z+\Delta z) = \frac{I_p(x,y,z)+\dfrac{\lambda_s}{\lambda_p}I_z(x,y,z)}{\dfrac{\lambda_s}{\lambda_p}+\dfrac{I_p(x,y,z)}{I_s(x,y,z)}\exp\left\{-\left[I_p(x,y,x)+\dfrac{\lambda_s}{\lambda_p}I_s(x,y,z)\right]g\Delta z\right\}}, \tag{9.102}$$

$$I_p(x,y,z+\Delta z) = I_p(x,y,z)+\frac{\lambda_s}{\lambda_p}[I_s(x,y,z)-I_s(x,y,z+\Delta z)]. \tag{9.103}$$

These equations are solved simultaneously along with diffraction propagation by the command, `raman`.

9.6.2 Raman Interactions with Multiple Stokes Beams

It is straightforward to extend Eq. (9.98) to include 1st and 2nd Stokes beams:

$$\frac{\partial P_j}{\partial z} = -\frac{g_{ps}}{2}S_j(\vec{S}^*\cdot\vec{P}), \tag{9.104}$$

$$\frac{\partial S_j}{\partial z} = \frac{g_{ps}}{2}P_j(\vec{P}^*\cdot\vec{S})-\frac{g_{st}}{2}T_j(\vec{T}^*\cdot\vec{S}),$$

$$\frac{\partial T_j}{\partial z} = -\frac{g_{st}}{2}S_j(\vec{S}^*\cdot\vec{T}),$$

where T is the 2nd Stokes amplitude, g_{ps} is the Raman gain for pump to 1st Stokes, and g_{st} is the Raman gain for 1st Stokes to 2nd Stokes.

In the case of a single longitudinal mode, Eq. (9.104) simplifies to

$$\frac{\partial P}{\partial z} = -\frac{g_{ps}}{2}|S|^2P, \tag{9.105}$$

$$\frac{\partial S}{\partial z} = \frac{g_{ps}}{2}|P|^2S-\frac{g_{st}}{2}|T|^2S,$$

$$\frac{\partial T}{\partial z} = -\frac{g_{st}}{2}|S|^2T.$$

Equation (9.105) is appropriate for single mode, multiple Stokes calculations.

9.6.3 Four-Wave Mixing

In the general case, four-wave mixing takes the form

$$\frac{\partial E_4}{\partial z} \propto \chi_{123} E_1 E_2 E_3 \exp[j(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3 - \boldsymbol{k}_4) \cdot \boldsymbol{z} - j(\omega_1 + \omega_2 + \omega_3 - \omega_4)t], \tag{9.106}$$

where ω_i and $\boldsymbol{k}_i$ are the angular frequency and wave number of the ith beam. The sign of change ω_i and $\boldsymbol{k}_i$ according to whether E or its complex conjugate is used in Equation (9.105). The term $\exp(j\boldsymbol{k} \cdot \boldsymbol{z})$ is the phase matching factor. In general, there is significant interaction only if both of the following conditions are met

$$\omega_1 + \omega_2 + \omega_3 - \omega_4 = 0, \tag{9.107}$$

$$(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3 - \boldsymbol{k}_4) \cdot \boldsymbol{z} = 0. \tag{9.108}$$

The latter term is often called the phase matching condition. As Eq. (9.105) is integrated over distance Δz, the phase matching condition determines how coherent the effect is. If $\boldsymbol{\Delta k} \cdot \boldsymbol{\Delta z} \gg 0$, the average gain goes to zero because of the rapidly rotating phase factor. In the general case, the normalized phase factor is,

$$\frac{1}{\Delta z}\int_0^{\Delta z} e^{j\Delta k \Delta z} dz = e^{j\Delta k \Delta z/2} \frac{\sin(\Delta k \Delta z/2)}{\Delta k \Delta z/2}. \tag{9.109}$$

The $\sin(x)/x$ function drives the effective gain to zero for large values of Δz.

9.6.3.1 Raman amplification as four-wave mixing

One may consider the single line Raman process to be a partially degenerate form of four-wave mixing in which

$$\omega_1 = \omega_s, \ \omega_2 = -\omega_s, \ \omega_3 = \omega_p, \text{ and } \omega_4 = \omega_p, \tag{9.110}$$

so that the exponential term for ω is exactly zero and the exponential term for k is zero in the absence of dispersion. Consequently there is no phase matching condition for Raman scattering.

9.6.3.2 General four-wave mixing equations

Four-wave mixing may cause 2nd Stokes production in a Raman cell when both pump and 1st Stokes beams are present in significant magnitude. The equations for four-wave mixing, including multiple longitudinal modes, is

$$\frac{\partial P_j}{\partial z} = -\frac{g_{4w}}{2} e^{i\boldsymbol{\Delta k} \cdot \boldsymbol{z}} S_j (\vec{\hat{T}}^* \cdot \vec{\hat{S}}), \tag{9.111}$$

$$\frac{\partial S_j}{\partial z} = -\frac{g_{4w}}{2} e^{-i\boldsymbol{\Delta k} \cdot \boldsymbol{z}} T_j (\vec{\hat{S}}^* \cdot \vec{\hat{P}}) + \frac{g_{4w}}{2} e^{-i\boldsymbol{\Delta k} \cdot \boldsymbol{z}} P_j (\vec{\hat{S}}^* \cdot \vec{\hat{T}}),$$

$$\frac{\partial T_j}{\partial z} = \frac{g_{4w}}{2} e^{i\Delta \boldsymbol{k} \cdot z} S_j(\vec{P}^* \cdot \vec{S}) .$$

where $\Delta \boldsymbol{k} = \boldsymbol{k}_p + \boldsymbol{k}_t - 2\boldsymbol{k}_s$ and g_{4w} is the gain coefficient for four-wave mixing.

The single mode form of Eq. (9.111) is shown below.

$$\frac{\partial P}{\partial z} = -\frac{g_{4w}}{2} e^{i\Delta \boldsymbol{k} \cdot z} S^2 T^* , \tag{9.112}$$

$$\frac{\partial S}{\partial z} = -\frac{g_{4w}}{2} e^{-i\Delta \boldsymbol{k} \cdot z} T S^* P + \frac{g_{4w}}{2} e^{-i\Delta \boldsymbol{k} \cdot z} P S^* T = 0 ,$$

$$\frac{\partial T}{\partial z} = \frac{g_{4w}}{2} e^{i\Delta \boldsymbol{k} \cdot z} S^2 P^* .$$

Note that $\partial S / \partial z = 0$, in the single mode case.

A closed form solution exists if no Raman amplification is considered and if the phase matching factor is similar to 1. Let,

$$\alpha = \langle \frac{g_{4w}}{2} e^{i\Delta \boldsymbol{k} \cdot z} S^2 \rangle , \tag{9.113}$$

where the average is taken over the propagation distance. The closed form solution is

$$I_p(z) = I_p(z) \cos(|\alpha| z) , \tag{9.114}$$

$$I_s(z) = \frac{\alpha}{|\alpha|} \frac{\lambda_p}{\lambda_s} I_p(0) \sin(|\alpha| z) .$$

9.6.3.3 Effect of dispersion

In the absence of dispersion $\Delta k = 0$, if all beams are collinear. In fact dispersion in typical gasses such as hydrogen and methane are significant. In the case of finite dispersion there are combinations of propagation directions that will result in $\Delta \boldsymbol{k} \cdot \hat{\boldsymbol{z}} = 0$, where $\boldsymbol{z}$ is the z-direction unit vector. The dispersion may be characterized by the Cauchy expression for index of refraction,

$$n = 1 + a_c P_{atm} \left(1 + \frac{b_c}{\lambda^2} + \ldots \right), \tag{9.115}$$

where a_c and b_c are the Cauchy coefficients and P_{atm} is the pressure in Amagats.

The wave number vector is

$$\boldsymbol{k} = \frac{2\pi}{\lambda}\left[1 + a_c P_{atm}\left(1 + \frac{b_c}{\lambda^2} + \ldots\right)\right]\hat{\boldsymbol{k}}. \tag{9.116}$$

The phase matching factor is $\Delta k = \boldsymbol{\Delta k} \cdot \hat{\boldsymbol{z}}$

$$\Delta k = 2\pi\left[f(\lambda_p)\frac{\hat{\boldsymbol{k}}_p \cdot \hat{\boldsymbol{z}}}{\lambda_p} + f(\lambda_t)\frac{\hat{\boldsymbol{k}}_t \cdot \hat{\boldsymbol{z}}}{\lambda_t} + f(\lambda_s)\frac{\hat{\boldsymbol{k}}_s \cdot \hat{\boldsymbol{z}}}{\lambda_s}\right], \tag{9.117}$$

where the function $f(\lambda)$ is defined to be

$$f(\lambda) = 1 + a_c P_{atm}\left(1 + \frac{b_c}{\lambda^2} + \ldots\right). \tag{9.118}$$

Because the refractive index is similar to one, certain approximations are usually made. We separate the expression Δk into its dispersive Δk_{disp} and nondispersive parts Δk_{angle}. The nondispersive term can only vary with angle. The dispersive term varies with angle and with change in index of refraction

$$\Delta k = \Delta k_{angle} + \Delta k_{disp}, \tag{9.119}$$

$$\Delta k_{angle} = 2\pi(1 + a_c P_{atm})\left[\frac{\hat{\boldsymbol{k}}_p \cdot \hat{\boldsymbol{z}}}{\lambda_p} + \frac{\hat{\boldsymbol{k}}_t \cdot \hat{\boldsymbol{z}}}{\lambda_t} - 2\frac{\hat{\boldsymbol{k}}_s \cdot \hat{\boldsymbol{z}}}{\lambda_s}\right], \tag{9.120}$$

$$\Delta k_{disp} = 2\pi a_c b_c P_{atm}\left[\frac{\hat{\boldsymbol{k}}_p \cdot \hat{\boldsymbol{z}}}{\lambda_p} + \frac{\hat{\boldsymbol{k}}_t \cdot \hat{\boldsymbol{z}}}{\lambda_t} - 2\frac{\hat{\boldsymbol{k}}_s \cdot \hat{\boldsymbol{z}}}{\lambda_s}\right]. \tag{9.121}$$

In the expression for Δk_{angle}, $a_c P_{atm}$ can be neglected. The small angle approximation can also be made for $\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{z}}$,

$$\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{z}} \cong 1 - \frac{\theta^2}{2}, \tag{9.122}$$

where θ is the angle between $\boldsymbol{k}$ and $\boldsymbol{z}$. Also, it should be noted that if λ_p, λ_s, and λ_t satisfy the Raman conditions,

$$\omega_s = \omega_p - \Delta\omega_R, \tag{9.123}$$

$$\omega_t = \omega_p - 2\Delta\omega_R,$$

where $\Delta\omega_R$ is the Raman frequency shift then,

$$\frac{1}{\lambda_p^3} + \frac{1}{\lambda_t^3} - \frac{2}{\lambda_s^3} = 0. \tag{9.124}$$

The change in angle is, therefore,

$$\Delta k_{angle} = -\pi\left[\frac{\theta_p^3}{\lambda_p^3} + \frac{\theta_t^3}{\lambda_t^3} - 2\frac{\theta_s^3}{\lambda_s^3}\right]. \tag{9.125}$$

In the dispersive term, the angular dependency may be neglected, $\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{z}} \cong 0$,

$$\Delta k_{disp} \cong 2\pi a_c b_c P_{atm}\left[\frac{1}{\lambda_p} + \frac{1}{\lambda_t} - 2\frac{1}{\lambda_s}\right]. \tag{9.126}$$

The final expression is obtained from Eqs. (9.125) and (9.126)

$$\Delta k \cong 2\pi a_c b_c P_{atm}\left[\frac{1}{\lambda_p} + \frac{1}{\lambda_t} - 2\frac{1}{\lambda_s}\right] - \pi\left[\frac{\theta_p^3}{\lambda_p^3} + \frac{\theta_t^3}{\lambda_t^3} - 2\frac{\theta_s^3}{\lambda_s^3}\right]. \tag{9.127}$$

9.6.3.4 Sample calculation: dispersion in four wave mixing

To illustrate these effects, consider the case of an excimer/Raman system with

$$\lambda_p = 0.353 \text{ micron}, \ \lambda_s = 0.414 \text{ micron}, \ \lambda_t = 0.499 \text{ micron}.$$

In this case the pump and 2nd Stokes beam are to move parallel to the optical axis and the 1st Stokes beam is at an angle, θ_s. Assume a pressure P_{atm} Patm = 3. For hydrogen a_c = 13.6 x 10-5, b_c = 7.70 x 10-11. Eqs. (9.125) and (9.126) take the value

$$\Delta k_{angle} = -\pi\left[\frac{\theta_p^3}{\lambda_p^3} + \frac{\theta_t^3}{\lambda_t^3} - 2\frac{\theta_s^3}{\lambda_s^3}\right] = -1.52\times10^5\theta_s^2 \text{ cm}^{-1},$$

$$\Delta k_{disp} \cong 2\pi a_c b_c P_{atm}\left[\frac{1}{\lambda_p} + \frac{1}{\lambda_t} - 2\frac{1}{\lambda_s}\right] = 0.51 \text{ cm}^{-1},$$

$$\Delta k = 0.51 - 1.52\times10^5\theta_s^2 \text{ cm}^{-1}.$$

If $\theta_s = 0$, then a propagation of about $\Delta z = 12.3$ cm will result in a value of $\Delta k \Delta z = 2\pi$, which drives four wave mixing to zero. $\Delta k \Delta z = 0$ if $\theta_s \cong 0.0018$ radians. This is the phase matching condition which results in maximum four-wave mixing amplification.

9.6.3.5 Calculation of the wavefront angles

Equation (9.127) requires determination of the propagation angles of the *P*, *S*, and *T* beams with respect to the z-axis. In the case of beams with significant wavefront curvature or aberration the beam angles vary across the aperture. In this case the local angles are used. The local angles are calculated from the complex amplitude, as follows. The value for θ_x will be derived here. The derivation for θ_y is identical. Let $a(i+1)$ and $a(i-1)$ be the complex amplitude of two points adjacent to $a(i)$. The difference complex amplitude is

$$\Delta a = a(i+q) - a(i-1). \tag{9.128}$$

The unit vector (in complex space) parallel to $a(i)$ is $\hat{a}$:

$$\hat{a} = \frac{a(i)}{|a(i)|}. \tag{9.129}$$

The component of Δa perpendicular to $\hat{a}$ is

$$\Delta a_p = \Delta a - \mathrm{Re}[\Delta a \hat{a}]\hat{a}. \tag{9.130}$$

The local wavefront angle is

$$\theta_{\text{local}} = \frac{1}{2}\frac{|\Delta a_p|}{|a(i)|}. \tag{9.131}$$

In addition we need to consider the overall tilt direction of the beam, θ_{tilt}, and the radius of curvature, *R*. The total wavefront angle is

$$\theta_{\text{total}} = \theta_{\text{local}} + \frac{x}{R} + \theta_{\text{tilt}}. \tag{9.132}$$

A similar calculation is used to get the y-angles. The total angle is used in Eq. (9.122) to determine the phase matching condition.

9.7 Optical Parametric Oscillator

For three plane waves, the steady-state equations for the optical parametric oscillator (OPO) is:

$$\frac{\partial}{\partial z}E_p(\boldsymbol{k}_p, z) = i\kappa_p E_s(\boldsymbol{k}_s, z)E_i(\boldsymbol{k}_i, z)e^{-i\Delta \boldsymbol{k}\cdot \boldsymbol{z}}, \tag{9.133}$$

$$\frac{\partial}{\partial z}E_s(\boldsymbol{k}_s, z) = i\kappa_s^* E_p(\boldsymbol{k}_p, z)E_i^*(\boldsymbol{k}_i, z)e^{i\Delta\boldsymbol{k}\cdot z}, \tag{9.134}$$

$$\frac{\partial}{\partial z}E_i(\boldsymbol{k}_i, z) = i\kappa_i E_p(\boldsymbol{k}_p, z)E_s^*(\boldsymbol{k}_s, z)e^{i\Delta\boldsymbol{k}\cdot z}, \tag{9.135}$$

$$\Delta\boldsymbol{k} = \boldsymbol{k}_p - \boldsymbol{k}_s - \boldsymbol{k}_i. \tag{9.136}$$

Direct solution of these equations for beams containing a distribution of k-vectors (true of beams with aberration and/or finite irradiance distributions) requires that for each k-vector component to be integrated in z double integral over k-vectors be evaluated:

$$\frac{\partial}{\partial z}E_p(\boldsymbol{k}_p, z) = i\kappa_p \int_{\boldsymbol{k}_s} d\boldsymbol{k}_s \int_{\boldsymbol{k}_i} d\boldsymbol{k}_i \tilde{E}_s^*(\boldsymbol{k}_s, z)\tilde{E}_i(\boldsymbol{k}_i, z)e^{-i\Delta\boldsymbol{k}\cdot z}, \tag{9.137}$$

$$\frac{\partial}{\partial z}E_s(\boldsymbol{k}_s, z) = i\kappa_s^* \int_{\boldsymbol{k}_p} d\boldsymbol{k}_p \int_{\boldsymbol{k}_i} d\boldsymbol{k}_i \tilde{E}_p^*(\boldsymbol{k}_p, z)\tilde{E}_i^*(\boldsymbol{k}_i, z)e^{i\Delta\boldsymbol{k}\cdot z}, \tag{9.138}$$

$$\frac{\partial}{\partial z}E_i(\boldsymbol{k}_i, z) = i\kappa_i^* \int_{\boldsymbol{k}_p} d\boldsymbol{k}_p \int_{\boldsymbol{k}_s} d\boldsymbol{k}_s \tilde{E}_p(\boldsymbol{k}_p, z)\tilde{E}_s^*(\boldsymbol{k}_s, z)e^{i\Delta\boldsymbol{k}\cdot z}. \tag{9.139}$$

Solving the equations as defined in Eqs. (9.137)–(9.139) is difficult because there is a double integral for each of three fields at each point to be evaluated. A direct solution of Eqs. (9.137)–(9.139) is quite impractical for numerical evaluation of the transverse modes in a laser cavity where arrays of typical size of 128 x 128 must be propagated 100's of times to reach the steady-state condition — on the order of 10^6 double integrals must be evaluated.

9.7.1 Formulation for Efficient Solution

We can rewrite the equations in a form which will execute efficiently. The key is to perform the kinetics calculations in the spatial domain and intersperse diffraction propagation steps. Let us write each field in the form $E(x, y;z)\exp(j(\boldsymbol{k}\cdot\boldsymbol{r}))$, where $\boldsymbol{k}$ is a constant vector representing the average k-vector for the beam. The residual angular errors manifest themselves as phase errors in the three complex amplitudes transverse distributions. By including steps of diffraction propagation, the three transverse fields will exhibit the correct diffraction behavior.

Propagation of a plane wave is described in Section 3.5. From Eq. (3.36) we have a factorization of the phase due to propagation into $\exp(jkz)$ and $\exp[-(jkz/2)(\alpha^2+\beta^2)]$ in OPO calculations the constant phase. The effect of $\exp(jkz)$ will be included in the solution of the differential equations in the spatial domain and $\exp[-(jkz/2)(\alpha^2+\beta^2)]$, the quadratic phase factor, and will be included, as usual, by the standard diffraction propagation routines in GLAD.

Smith, et. al., used a similar procedure with the principle distinctions that they explicitly form the three medium polarizations based on the products $P_p \propto E_s E_i$, $P_s \propto E_p E_i^*$, and $P_i \propto E_p E_s^*$ and then Fourier

transform the medium polarizations to perform the kinetic calculations in the frequency domain[8]. This approach requires Fourier transformation of six fields: E_p, E_s, E_i, P_p, P_s, and P_i. For the model used in GLAD, the medium effects are implemented in the spatial domain so that only the optical fields need be Fourier transformed, cutting the number of Fourier transform operations from six to three per step. The step length requirement is the same for both methods. Smith, et. al., also integrate the three kinetic equations directly where the model in GLAD uses a closed-form solution of the paired signal and idler equations and incorporate this closed-form solution into a split-step solution with the kinetic equation for the pump, resulting in excellent numerical stability for high gain cases. The set of differential equations is

We may write the diffraction effects as differential equations

$$\frac{\partial}{\partial z}E_p(x,y,z) = i\kappa_p E_s(x,y,z)E_i(x,y,z)e^{-i(k_p-k_s-k_i)z}, \tag{9.140}$$

$$\frac{\partial}{\partial z}E_s(x,y,z) = i\kappa_s^* E_p(x,y,z)E_i^*(x,y,z)e^{i(k_p-k_s-k_i)z}, \tag{9.141}$$

$$\frac{\partial}{\partial z}E_i(x,y,z) = i\kappa_i E_p(x,y,z)E_s^*(x,y,z)e^{i(k_p-k_s-k_i)z}. \tag{9.142}$$

We may write the diffraction effects as differential equations

$$\frac{\partial}{\partial z}E_p(x,y,z) = -\frac{i}{2k_p}\nabla_\perp^2 E_p(x,y,z), \tag{9.143}$$

$$\frac{\partial}{\partial z}E_s(x,y,z) = -\frac{i}{2k_s}\nabla_\perp^2 E_s(x,y,z), \tag{9.144}$$

$$\frac{\partial}{\partial z}E_i(x,y,z) = -\frac{i}{2k_i}\nabla_\perp^2 E_i(x,y,z), \tag{9.145}$$

but these are conveniently solved by Fourier transforms using FFT methods according to the usual propagation routines in GLAD.

The kinetics equations my be solved by finite difference methods. Integrating Eqs. (9.140)–(9.142) from 0 to Δz, we have

$$\Delta E_p(x,y,z) = i\kappa_p E_s(x,y,z)E_i(x,y,z)\int_0^{\Delta z} e^{-i\Delta kz}dz, \tag{9.146}$$

$$\Delta E_s(x,y,z) = i\kappa_s^* E_p(x,y,z)E_i^*(x,y,z)\int_0^{\Delta z} e^{i\Delta kz}dz, \tag{9.147}$$

$$\Delta E_i(x,y,z) = i\kappa_i^* E_p(x,y,z)E_s^*(x,y,z)\int_0^{\Delta z} e^{i\Delta kz}dz\,, \tag{9.148}$$

where the scalar dispersion term is

$$\Delta k = k_p - k_s - k_i\,. \tag{9.149}$$

The dispersion terms of Eqs. (9.146)–(9.148) may be integrated according to Eq. (9.109). The sin(x)/x function drives the effective gain to zero for large values of $\Delta k\Delta z$ and goes to unity for small values,

$$\Delta E_p(x,y,z) = i\kappa_p E_s(x,y,z)E_i(x,y,z)\Delta z e^{-i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}\,, \tag{9.150}$$

$$\Delta E_s(x,y,z) = i\kappa_s^* E_p(x,y,z)E_i^*(x,y,z)\Delta z e^{i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}\,, \tag{9.151}$$

$$\Delta E_i(x,y,z) = i\kappa_i^* E_p(x,y,z)E_s^*(x,y,z)\Delta z e^{i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}\,. \tag{9.152}$$

Equations (9.142)–(9.144) and Eqs. (9.145)–(9.147) may be solved by split-step methods. Note that the double integrals of Eqs. (9.137)-(9.139) are not required.

The coupling coefficients may be simplified by assuming κ is real. Only one coupling constant is needed provided the optical frequency and wavenumber are included. In gaussian units, according to Shen[9]:

$$\Delta E_p(x,y,z) = i\frac{\omega_p^2}{k_p}\kappa E_s(x,y,z)E_i(x,y,z)\Delta z e^{-i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}\,, \tag{9.153}$$

$$\Delta E_s(x,y,z) = i\frac{\omega_s^2}{k_s}\kappa^* E_p(x,y,z)E_i^*(x,y,z)\Delta z e^{i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}\,, \tag{9.154}$$

$$\Delta E_i(x,y,z) = i\frac{\omega_i^2}{k_i}\kappa^* E_p(x,y,z)E_s^*(x,y,z)\Delta z e^{i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}\,. \tag{9.155}$$

9.7.2 Check of Energy Conservation

Energy is conserved, to first order, by the three coupled differential equations as well as the finite difference formulation. We use the fact that the irradiance is related to the electric field by $I = |E|^2 cn\varepsilon_0/2$. The electric permeability of free space ε_0 is defined to be $\varepsilon_0 = 10^7/4\pi c$ faraday/m or $\varepsilon_0 = 10^5/4\pi c$ faraday/cm (we also have the relationship $c = 1/\sqrt{\mu_0\varepsilon_0}$), where $\mu_0 = 4\pi\times10^{-10}$ henry/m. For an incremental change in electric field, the change in irradiance is $\Delta I = cn\varepsilon_0\mathrm{Re}[E^*E] = cn\varepsilon_0\mathrm{Re}[EE^*]$:

$$\Delta I_p = -2\mathrm{Re}\left[\frac{cn_p\varepsilon_0}{2}E_p\Delta E_p^*\right] = -\omega_p\mathrm{Re}\left[ic^2\varepsilon_0 E_p E_s^* E_i^*\int_0^{\Delta z} e^{i\Delta kz}dz\right], \tag{9.156}$$

$$\Delta I_s = 2\mathrm{Re}\left[\frac{cn_s\varepsilon_0}{2}E_s^*\Delta E_p\right] = \omega_s\mathrm{Re}\left[ic^2\varepsilon_0\kappa E_p E_s^* E_i^*\int_0^{\Delta z} e^{i\Delta kz}dz\right], \tag{9.157}$$

$$\Delta I_i = 2\mathrm{Re}\left[\frac{cn_s\varepsilon_0}{2}E_i^*\Delta E_i\right] = \omega_i\mathrm{Re}\left[ic^2\varepsilon_0\kappa E_p E_s^* E_i^*\int_0^{\Delta z} e^{i\Delta kz}dz\right], \tag{9.158}$$

where we take advantage of the frequency matching condition $\omega_p - \omega_s - \omega_i = 0$. The finite difference equations conserve energy to first order, but energy is not conserved to second order by the finite difference equations.

9.7.2.1 Expression in MKSA units

From Eq. (9.153–9.155) we have

$$\Delta E_p(x,y,z) = i\frac{c\omega_p}{n_p}\kappa E_s(x,y,z)E_i(x,y,z)\Delta z e^{-i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}, \tag{9.159}$$

$$\Delta E_s(x,y,z) = i\frac{c\omega_s}{n_s}\kappa^* E_p(x,y,z)E_i^*(x,y,z)\Delta z e^{i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}, \tag{9.160}$$

$$\Delta E_i(x,y,z) = i\frac{c\omega_i}{n_i}\kappa^* E_p(x,y,z)E_s^*(x,y,z)\Delta z e^{i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}. \tag{9.161}$$

We can write κ in terms of the nonlinear susceptibility

$$\kappa = \frac{2\pi}{c^2}\chi_{\mathrm{esu}}^{(2)}, \tag{9.162}$$

where $\chi_{\mathrm{esu}}^{(2)}$ is in gaussian units:

$$\Delta E_p(x,y,z) = i\frac{\omega_p}{n_p}\frac{2\pi}{c^2}\chi_{\mathrm{esu}}^{(2)}E_s(x,y,z)E_i(x,y,z)\Delta z e^{-i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}, \tag{9.163}$$

$$\Delta E_s(x,y,z) = i\frac{\omega_s}{n_s}\frac{2\pi}{c^2}\chi_{\mathrm{esu}}^{(2)*}E_p(x,y,z)E_i^*(x,y,z)\Delta z e^{i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}, \tag{9.164}$$

$$\Delta E_i(x,y,z) = i\frac{\omega_i}{n_i}\frac{2\pi}{c^2}\chi^{(2)*}_{\text{esu}}E_p(x,y,z)E_s^*(x,y,z)\Delta z e^{i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}. \tag{9.165}$$

The above equations are in gausian units following Shen. It is more convenient to use MKSA than gaussian units (GLAD is MKSA with the exception that distance is in centimeters). Converting to MKSA we employ the following substitutions

$$\chi^{(2)}_{\text{esu}} \to \frac{3\times10^4}{4\pi\varepsilon_0}\chi^{(2)}_{\text{MKSA}},\ c \to \frac{1}{\sqrt{\mu_0\varepsilon_0}},\ E_{\text{esu}} \to \frac{1}{3\times10^4}E_{\text{MKSA}}. \tag{9.166}$$

From this point forward we will refer to $\chi^{(2)}_{\text{MKSA}}$ as simply $\chi^{(2)}$. $\chi^{(2)}$ is twice the d-coefficient value which is often seen in the literature and has units of meters/volt.

Rewriting Eqs. (9.163)–(9.165) entirely in MKSA, we have

$$\Delta E_p(x,y,z) = \frac{i}{2}\frac{\omega_p}{n_p}\chi^{(2)}\left(\frac{\mu_0}{\varepsilon_0}\right)^{1/2}E_s(x,y,z)E_i(x,y,z)\Delta z e^{-i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}, \tag{9.167}$$

$$\Delta E_s(x,y,z) = \frac{i}{2}\frac{\omega_s}{n_s}\chi^{(2)*}\left(\frac{\mu_0}{\varepsilon_0}\right)^{1/2}E_p(x,y,z)E_i^*(x,y,z)\Delta z e^{i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}, \tag{9.168}$$

$$\Delta E_i(x,y,z) = \frac{i}{2}\frac{\omega_i}{n_i}\chi^{(2)*}\left(\frac{\mu_0}{\varepsilon_0}\right)^{1/2}E_p(x,y,z)E_s^*(x,y,z)\Delta z e^{i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}. \tag{9.169}$$

In terms of the complex amplitude in GLAD

$$|A|^2 = I = \frac{cn\varepsilon_0}{2}|E|^2 \text{ and } A \equiv E\sqrt{\frac{n}{2\mu}}\sqrt{\frac{\varepsilon_0}{\mu_0}} = E\sqrt{\frac{n}{2}}\sqrt{\frac{\varepsilon_0}{\mu_0}}, \tag{9.170}$$

where $\mu = 1$ for ordinary refractive materials.

The phase of the three fields is calculated and stored in double precision to allow for longer propagation distances and incorporated into the term $\exp(i\Delta\phi)$ which includes the absolute phase of the three fields $\Delta\phi = \phi_p - \phi_s - \phi_i$. The coupling constant may be conveniently be redefined:

$$\tilde{\kappa} = \frac{1}{\sqrt{2}}\omega_s\left(\frac{\mu_0}{\varepsilon_0}\right)^{3/4}\sqrt{\frac{1}{n_p n_s n_i}}\chi^{(2)}e^{-i\Delta\varphi} = 9.6668517\times10^{-5}\omega_s\sqrt{\frac{1}{n_p n_s n_i}}\chi^{(2)}e^{-i\Delta\varphi}. \tag{9.171}$$

$\tilde{\kappa}$ has units of [$w^{-1/2}$], we may calculate gain in terms of pump irradiance and $\tilde{\kappa}$

$$\Delta A_p(x,y,z) = i\frac{\omega_p}{\omega_s}\tilde{\kappa}A_s(x,y,z)A_i(x,y,z)\Delta z e^{-i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}, \tag{9.172}$$

$$\Delta A_s(x,y,z) = i\tilde{\kappa}^* A_p(x,y,z)A_i^*(x,y,z)\Delta z e^{i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}, \tag{9.173}$$

$$\Delta A_i(x,y,z) = i\frac{\omega_i}{\omega_s}\tilde{\kappa}^* A_p(x,y,z)A_s^*(x,y,z)\Delta z e^{i\Delta kz/2}\frac{\sin(\Delta kz/2)}{\Delta kz/2}. \tag{9.174}$$

The characteristic lengths for these equations are

$$z_{\text{crit}_p} = \frac{1}{|\tilde{\kappa}|\frac{\omega_p}{\omega_s}|A_s||A_i|}, z_{\text{crit}_s} = \frac{1}{|\tilde{\kappa}||A_p||A_i|}, z_{\text{crit}_i} = \frac{1}{|\tilde{\kappa}|\frac{\omega_i}{\omega_s}|A_p||A_i|}, z_{\text{crit}_{\Delta k}} = \frac{1}{\Delta k}. \tag{9.175}$$

The least of the critical lengths should be used in solving Eqs. (9.172)–(9.174) should be selected and the integration step length set well below the minimum of the critical lengths.

In the special case where $|A_p| \gg |A_s| + |A_i|$ then

$$\left|\frac{\partial A_p}{\partial z}\right| \ll \left|\frac{\partial A_s}{\partial z}\right| + \left|\frac{\partial A_i}{\partial z}\right|, \tag{9.176}$$

and we may solve Eqs. (9.172) and (99.174) assuming A_p is effectively constant

$$\frac{\partial A_s}{\partial z} = i\tilde{\kappa}^* A_p A_i^* e^{i\Delta kz}, \tag{9.177}$$

$$\frac{\partial A_i^*}{\partial z} = -i\frac{\omega_i}{\omega_s}\tilde{\kappa}A_p^* A_s e^{-i\Delta kz}. \tag{9.178}$$

Equations (9.177) and (9.178) have a closed form solution

$$g_0^2 = 4\frac{\omega_i}{\omega_s}|\tilde{\kappa}|^2|A_p| = 2\omega_s\omega_i\left(\frac{\mu_0}{\varepsilon_0}\right)^{3/2}\frac{1}{n_p n_s n_i}|\chi^{(2)}|^2|A_p|^2 = \frac{\omega_s\omega_i\mu_0}{n_s n_i \varepsilon_0}|\chi^{(2)}|^2|E_p|^2, \tag{9.179}$$

$$g = \left(g_0^2 - \Delta k^2\right)^{1/2}, \tag{9.180}$$

$$A_s(z) = (C_{s+}e^{-gz/2} + C_{s-}e^{gz/2})e^{i\Delta kz/2}, \tag{9.181}$$

$$A_i^*(z) = (C_{i+}e^{-gz/2} + C_{i-}e^{gz/2})e^{-i\Delta kz/2}. \tag{9.182}$$

We use the results of the closed form solution of A_s and A_i to calculate ΔA_p

$$C_{s+} = \frac{1}{g}\left[\frac{g + i\Delta k}{2}A_s(0) - i\tilde{\kappa}^* A_p A_i^*(0)\right], \tag{9.183}$$

$$C_{s-} = \frac{1}{g}\left[\frac{g - i\Delta k}{2}A_s(0) + i\tilde{\kappa}^* A_p A_i^*(0)\right], \tag{9.184}$$

$$C_{i+} = \frac{1}{g}\left[\frac{g + i\Delta k}{2}A_i(0) - i\frac{\omega_i}{\omega_s}\tilde{\kappa}^* A_p A_s^*(0)\right], \tag{9.185}$$

$$C_{i-} = \frac{1}{g}\left[\frac{g - i\Delta k}{2}A_i(0) + i\frac{\omega_i}{\omega_s}\tilde{\kappa}^* A_p A_s^*(0)\right]. \tag{9.186}$$

Regrouping for better numerical solution

$$A_s(z) = \left\{A_s(0)\cosh\left(\frac{gz}{2}\right) + \left[-i\Delta kA_s(0) + i\sqrt{\frac{\omega_s}{\omega_i}}g_0A_i^*(0)\right]\left[\frac{1}{g}\sinh\left(\frac{gz}{2}\right)\right]\right\}e^{i\Delta kz/2}, \tag{9.187}$$

$$A_i(z) = \left\{A_i(0)\cosh\left(\frac{gz}{2}\right) + \left[-i\Delta kA_i(0) + i\sqrt{\frac{\omega_i}{\omega_s}}g_0A_s^*(0)\right]\left[\frac{1}{g}\sinh\left(\frac{gz}{2}\right)\right]\right\}e^{i\Delta kz/2}. \tag{9.188}$$

For small g we may eliminate the factor of $1/g$

$$A_s(z) = \left\{A_s(0)\cosh\left(\frac{gz}{2}\right) + \left[-i\Delta kA_s(0) + i\sqrt{\frac{\omega_s}{\omega_i}}g_0A_i^*(0)\right]\left[\frac{z}{2} + \frac{g^2}{3!}\left(\frac{z}{2}\right)^3 + \frac{g^4}{5!}\left(\frac{z}{2}\right)^5 + \ldots\right]\right\}e^{i\Delta kz/2}, \tag{9.189}$$

$$A_i(z) = \left\{A_i(0)\cosh\left(\frac{gz}{2}\right) + \left[-i\Delta kA_i(0) + i\sqrt{\frac{\omega_i}{\omega_s}}g_0A_s^*(0)\right]\left[\left[\frac{z}{2} + \frac{g^2}{3!}\left(\frac{z}{2}\right)^3 + \frac{g^4}{5!}\left(\frac{z}{2}\right)^5 + \ldots\right]\right]\right\}e^{i\Delta kz/2}. \tag{9.190}$$

We may write the condition $|\Delta k| > |g_0|$ in terms of $g = ig'$

$$A_s(z) = \left\{A_s(0)\cosh\left(\frac{gz}{2}\right) + \left[-i\Delta k A_s(0) + i\sqrt{\frac{\omega_s}{\omega_i}}g_0 A_i^*(0)\right]\left[\frac{1}{g}\sinh\left(\frac{g'z}{2}\right)\right]\right\}e^{i\Delta kz/2}, \tag{9.191a}$$

$$A_i(z) = \left\{A_i(0)\cosh\left(\frac{gz}{2}\right) + \left[-i\Delta k A_i(0) + i\sqrt{\frac{\omega_i}{\omega_s}}g_0 A_s^*(0)\right]\left[\frac{1}{g}\sinh\left(\frac{g'z}{2}\right)\right]\right\}e^{i\Delta kz/2}. \tag{9.192}$$

For both n $|\Delta k| > |g_0|$ and small g, we may eliminate the factor of $1/g$

$$A_s(z)\big|_{g'z \ll 1} = \left\{A_s(0)\cosh\left(\frac{gz}{2}\right) + \left[-i\Delta k A_s(0) + i\sqrt{\frac{\omega_s}{\omega_i}}g_0 A_i^*(0)\right]\frac{z}{2}\right\}e^{i\Delta kz/2}, \tag{9.193}$$

$$A_i(z)\big|_{g'z \ll 1} = \left\{A_i(0)\cosh\left(\frac{gz}{2}\right) + \left[-i\Delta k A_i(0) + i\sqrt{\frac{\omega_i}{\omega_s}}g_0 A_s^*(0)\right]\frac{z}{2}\right\}e^{i\Delta kz/2}. \tag{9.194}$$

We use the results of the closed form solution of A_s and A_i to calculate ΔA_p.

$$\Delta A_p = i\frac{\omega_p}{\omega s}\tilde{\kappa}\int_0^{\Delta z} A_s(z)A_i(z)\ dz. \tag{9.195}$$

Solving the Eq. (9.195) for $|g_0| > |\Delta k|$, we have

$$A_s(z) = \left\{B_1\cosh\left(\frac{gz}{2}\right) + B_2\left[\frac{1}{g}\sinh\left(\frac{gz}{2}\right)\right]\right\}e^{i\Delta kz/2}, \tag{9.196}$$

$$A_i(z) = \left\{B_3\cosh\left(\frac{gz}{2}\right) + B_4\left[\frac{1}{g}\sinh\left(\frac{gz}{2}\right)\right]\right\}e^{i\Delta kz/2}, \tag{9.197}$$

$$B_1 = A_s(0),\ B_2 = -i\Delta k A_s(0) + i\sqrt{\frac{\omega_s}{\omega_i}}g_0 A_i^*(0), \tag{9.198}$$

$$B_3 = A_i(0),\ B_4 = -i\Delta k A_i(0) + i\sqrt{\frac{\omega_i}{\omega_s}} g_0 A_s^*(0),$$

$$\Delta A_p \quad , \tag{9.199}$$

$$= i\frac{\omega_p}{\omega s}\frac{\tilde{\kappa}}{2}\int_0^{\Delta z} B_1 B_3 \cosh^2\left(\frac{gz}{2}\right) + (B_2 B_3 + B_1 B_4)\frac{1}{g}\sinh\left(\frac{gz}{2}\right)\cosh\left(\frac{gz}{2}\right) + B_2 B_4 \frac{1}{g^2}\sinh^2\left(\frac{gz}{2}\right) dz$$

$$\Delta A_p = i\frac{\omega_p}{\omega s}\frac{\tilde{\kappa}}{2}\left\{B_1 B_3\left[\frac{1}{g}\sinh\left(\frac{g\Delta z}{2}\right) + \Delta z\right] + (B_2 B_3 + B_1 B_4)\frac{1}{g^2}[\cosh(g\Delta z) - 1]\right. \quad . \tag{9.200}$$
$$\left. + B_2 B_4 \frac{1}{g^2}\left[\frac{1}{g}\sinh\left(\frac{g\Delta z}{2}\right) - \Delta z\right]\right\}$$

Solving the Eq. (9.195) for large dispersion, $|\Delta k| > |g_0|$, we have

$$A_s(z) = \left\{B_1 \cos\left(\frac{g' z}{2}\right) + B_2\left[\frac{1}{g'}\sin\left(\frac{g' z}{2}\right)\right]\right\} e^{i\Delta k z/2}, \tag{9.201}$$

$$A_i(z) = \left\{B_3 \cosh\left(\frac{g' z}{2}\right) + B_4\left[\frac{1}{g'}\sinh\left(\frac{g' z}{2}\right)\right]\right\} e^{i\Delta k z/2}, \tag{9.202}$$

$$\Delta A_p = i\frac{\omega_p}{\omega s}\frac{\tilde{\kappa}}{2}\int_0^{\Delta z} B_1 B_3 \cos^2\left(\frac{g' z}{2}\right) + (B_2 B_3 + B_1 B_4)\frac{1}{g'}\sin\left(\frac{g' z}{2}\right)\cos\left(\frac{g' z}{2}\right) + B_2 B_4 \frac{1}{g'^2}\sin^2\left(\frac{g' z}{2}\right) dz, \tag{9.203}$$

$$\Delta A_p = i\frac{\omega_p}{\omega s}\frac{\tilde{\kappa}}{2}\left\{B_1 B_3\left[\frac{1}{g'}\sin(g'\Delta z) + \Delta z\right] + (B_2 B_3 + B_1 B_4)\frac{1}{g'^2}[1 - \cos(g'\Delta z)]\right. \quad . \tag{9.204}$$
$$\left. + B_2 B_4 \frac{1}{g'^2}\left[\Delta z - \frac{1}{g'}\sin(g'\Delta z)\right]\right\}$$

For small values of g'

$$\Delta A_p = i\frac{\omega_p}{\omega s}\frac{\tilde{\kappa}}{2}\left\{B_1 B_3\left[2\Delta z - \frac{g'^2 \Delta z^3}{3!}\right] + (B_2 B_3 + B_1 B_4)\frac{\Delta z^2}{2} + B_2 B_4 \frac{\Delta z^3}{3!}\right\}. \tag{9.205}$$

9.7.3 Spatial Frequency Content and Axial Sampling

The axial steps must be sufficiently short that the angles are well-resolved. The characteristic diffraction length associated with a given angle θ is

$$z_{char} = \frac{\lambda}{2\theta^2}. \tag{9.206}$$

For a given sample spacing of Δx the maximum angle is $\theta_{\max} = \lambda/\Delta x$,

$$z_{char} = \frac{\Delta x^2}{2\lambda}. \tag{9.207}$$

Equation (9.207) gives a value of characteristic diffraction length sufficient to satisfy any angle in the computer array with sample spacing of Δx. The actual axial step size should be chosen to be several times less than the characteristic length given by Eq. (9.207).

Consider a wavelength of 0.4 microns, and a maximum angle of 0.005 radian, the sample spacing would be 0.01 cm and the characteristic length would be 1 cm. A good value for axial sampling would be 0.1 or 0.05 cm, giving 10 to 20 steps/cm — a quite reasonable value.

9.7.3.1 Numerical example, MKSA and ESU

In the limit of weak pump depletion, the small signal gain may be calculated in terms of electric field. From Eq. (9.179):

$$g_0 = \sqrt{\frac{\omega_s \omega_i}{n_s n_i}}\sqrt{\frac{\mu_0}{\varepsilon_o}}\left|\chi^{(2)}_{\text{MKSA}}\right|\left|E_p\right|. \tag{9.208}$$

Shen gives a numerical example with $\chi_{\text{esu}} = 2.7\times10^{-8}$ esu, $\lambda_1 \approx \lambda_2 \approx 1.06\mu$, $n_1 = n_2 = 2.23$. We have the relationship

$$\chi^{(2)}_{\text{MKSA}} = \frac{4\pi\varepsilon_0}{3\times10^4}\chi^{(2)}_{\text{esu}} \tag{9.209}$$

giving

$$\chi^{(2)}_{\text{MKSA}} = 10.02\times10^{-23}. \tag{9.210}$$

$g_0 = 3.01\times10^{-5}E_p$. Taking $E_p = 100$ esu and $n_p \approx 2.2$, equivalent to

$$E_p = 100 \text{ esu} \rightarrow I_p = \frac{cn\varepsilon_0}{2}\left|100 \text{ esu}\times 3\times10^4\right|^2 = 2.63\times10^{10}\text{w/m}^2 = 2.63\times10^6\text{w/cm}^2. \tag{9.211}$$

9.8 Sum-Frequency Generation (SFG)

Sum-frequency generation (SFG) is used for up-conversion. See Yariv, Ref. 10. Parametric interactions in a crystal can be used to convert a low frequency ω_1 to a high frequency ω_3 by mixing it with a strong laser beam at ω_2. In a representative case a pump of 1.06 micron is mixed with a signal of 10.6 micron to create an up-converted to 0.96 micron light. Following the description by Yariv, a pump beam with frequency ω_1 and energy $\hbar\omega_1$, signal laser with frequency ω_2 and y $\hbar\omega_2$, and up-converted beam with ω_3 and $\hbar\omega_3$, we have the relationships:

$$\omega_1 + \omega_2 = \omega_3, \tag{9.212}$$

$$\text{net energy change} = \hbar\omega_3 - \hbar\omega_1 - \hbar\omega_2 = 0. \tag{9.213}$$

The up-converted frequency is the sum of the signal and laser pump frequencies. Equation (9.213) indicates that a photon of the signal and pump beams are annihilated to create a photon of the up-converted beam. The optical energy is conserved in the beams with no heat dissappated into the medium.

The fundamental differential equations are similar to the optical parameteric amplifier (OPA):

$$\frac{\partial}{\partial z}A_1(x,y,z) = i\frac{\omega_1}{\omega_2}\tilde{\kappa}^* A_2^*(x,y,z)A_3(x,y,z)e^{i\Delta kz}, \tag{9.214}$$

$$\frac{\partial}{\partial z}A_2(x,y,z) = i\tilde{\kappa}^* A_1^*(x,y,z)A_3(x,y,z)e^{i\Delta kz}, \tag{9.215}$$

$$\frac{\partial}{\partial z}A_3(x,y,z) = i\frac{\omega_3}{\omega_1}\tilde{\kappa}A_1(x,y,z)A_2(x,y,z)e^{-i\Delta kz}, \tag{9.216}$$

where

$$\kappa = \frac{1}{\sqrt{2}}\omega_2\left(\frac{\mu_0}{\varepsilon_0}\right)^{3/4}\frac{1}{\sqrt{n_1 n_2 n_3}}\chi^{(2)}. \tag{9.217}$$

We may consider A_1 to be the pump beam, A_2 the signal, and A_3 the up-converted signal. As the pump beam is usually weakly depleted, we may assume that it is relatively constant over short distances and reorganize the three coupled differential equations into the two coupled equations for the coupling of signal and up-converted beam, Eq. (9.215) and (9.216): The closed-form solution of these equation is:

$$A_2(z) = \left(C_{2+}e^{igz/2} + C_{2-}e^{-igz/2}\right)e^{i\Delta kz/2}, \tag{9.218}$$

$$A_3(z) = \left(C_{3+}e^{igz/2} + C_{3-}e^{-igz/2}\right)e^{-i\Delta kz/2}, \tag{9.219}$$

$$g_0^2 = g^2 - \Delta k^2 = 2\frac{\omega_2\omega_1}{n_1 n_2 n_3}\left(\frac{\mu_0}{\varepsilon_0}\right)^{3/2}\left|\chi^{(2)}\right|^2\left|A_1\right|^2 = \frac{\omega_2\omega_3}{n_2 n_3}\frac{\mu_0}{\varepsilon_0}\left|\chi^{(2)}\right|^2\left|E_1\right|^2, \tag{9.220}$$

with the rightmost expression written to be of the same from as is used by Yariv, Eq. (17.6-4), where our $\chi^{(2)}$ is the same as Yariv's d [10]. g is found by

$$g = +\sqrt{g_0^2 + \Delta k^2}. \tag{9.221}$$

The C-coefficients are found from:

$$C_{2+} = \frac{1}{2g}\left[(g - \Delta k)A_2(0) + \sqrt{\frac{\omega_2}{\omega_3}}g_0 A_3(0)\right], \tag{9.222}$$

$$C_{2-} = \frac{1}{2g}\left[(g + \Delta k)A_2(0) - \sqrt{\frac{\omega_2}{\omega_3}}g_0 A_3(0)\right], \tag{9.223}$$

$$C_{3+} = \frac{1}{2g}\left[(g + \Delta k)A_3(0) + \sqrt{\frac{\omega_3}{\omega_2}}g_0 A_2(0)\right], \tag{9.224}$$

$$C_{3-} = \frac{1}{2g}\left[(g - \Delta k)A_3(0) - \sqrt{\frac{\omega_3}{\omega_2}}g_0 A_2(0)\right]. \tag{9.225}$$

In the special case where $\Delta k = 0$,

$$g_0^2 = 2\frac{\omega_2\omega_1}{n_1 n_2 n_3}\left(\frac{\mu_0}{\varepsilon_0}\right)^{3/2}\left|\chi^{(2)}\right|^2\left|A_1\right|^2, \tag{9.226}$$

from Eq. (9.220)

$$A_2(z) = A_2(0)\cos(g_0 z/2) + i\sqrt{\frac{\omega_2}{\omega_3}}A_3(0)\sin(g_0 z/2), \tag{9.227}$$

$$A_3(z) = i\sqrt{\frac{\omega_3}{\omega_2}}A_2(0)\sin(g_0 z/2) + A_3(0)\cos(g_0 z/2). \tag{9.228}$$

For the simplest case of $\Delta k = 0$ and $A_3(0) = 0$:

$$A_2(z) = A_2(0)\cos(g_0 z/2), \tag{9.229}$$

$$A_3(z) = i\sqrt{\frac{\omega_3}{\omega_2}}A_2(0)\sin(g_0 z/2). \tag{9.230}$$

The irradiance values for this simple case are

$$I_2(z) = I_2(0)\cos^2(g_0 z/2), \tag{9.231}$$

$$I_3(z) = \frac{\omega_3}{\omega_2}I_2(0)\sin^2(g_0 z/2) \quad . \tag{9.232}$$

We use the results of the closed form solution of A_2 and A_3 to calculate ΔA_1,

$$\Delta A_1 = i\frac{\omega_1}{\omega_2}\int_0^{\Delta z} A_2(z)^* A_3(z)e^{i\Delta kz}dz, \tag{9.233}$$

$$\Delta A_1 = i\frac{\omega_1}{\omega_2}\left[C_{2+}^* C_{3-}\frac{1-e^{-g\Delta z}}{ig} + C_{2-}^* C_{3+}\frac{e^{g\Delta z}-1}{ig} + [C_{2+}^* C_{3+} + C_{2-}^* C_{3-}]\Delta z\right]. \tag{9.234}$$

For $g_0 > \Delta k$, g is real valued and we can replace the complex exponential treatment above with an expression in sines and cosines by regrouping terms in Eq. (9.218) and (9.219):

$$A_2(z) = [A_2(0)\cos(gz/2) + iC_2'\sin(gz/2)]e^{igz/2}, \tag{9.235}$$

$$A_3(z) = [A_3(0)\cos(gz/2) + iC_3'\sin(gz/2)]e^{-igz/2}, \tag{9.236}$$

$$C_2' = \frac{1}{g}\left[-\Delta k A_2(0) + \sqrt{\frac{\omega_2}{\omega_3}}g_0 A_3(0)\right], \tag{9.237}$$

$$C_3' = \frac{1}{g}\left[\Delta k A_3(0) + \sqrt{\frac{\omega_3}{\omega_2}}g_0 A_2(0)\right]. \tag{9.238}$$

Equation (9.234) can also be regrouped:

$$\Delta A_1 = \frac{\omega_1}{\omega_2}\frac{1}{2g}\tilde{\kappa}^*\left\{ i\left[A_2^*(0)A_3(0) - C_2^{'*}C_3^{'}\right]\sin(g\Delta z) + \left[C_2^{'*}A_3(0) - C_3^{'}A_2^*(0)\right]\left[1-\cos(g\Delta z)\right] \right. \quad . \quad (9.239)$$
$$\left. + i\left[A_2^*(0)A_3(0) + C_2^{'*}C_3^{'}\right]g\Delta z \right\}$$

9.8.1 Numerical Example, Calculation 1

It is desired to generate 0.96 μ light (λ_3) by mixing light from a 1.06 μ laser (λ_1) with light from a 10.6 μ laser (λ_2).

$I_1 = |A_1|^2 = 10^4 \text{w/cm}^2$, $I_2 = |A_2|^2 = 1\text{w/cm}^2$, $I_3 = |A_3|^2 = 0\text{w/cm}^2$, $\lambda_1 = 1.06\mu$, $\lambda_2 = 10.6\mu$, $\lambda_2 = 0.963636\mu$, $L = 1$ cm, $n_1 = n_2 = n_3 = 2.6$, $\chi^{(2)} = 1.1\times10^{-20}\ \text{C/V}^2$.

$\omega_1 = 1.777030\times10^{15}$, $\omega_1 = 1.777030\times10^{14}$, $\omega_2 = 1.954733\times10^{15}$

$$\kappa = \frac{1}{\sqrt{2}}\omega_2\left(\frac{\mu_0}{\varepsilon_0}\right)^{3/4}\frac{1}{\sqrt{n_1n_2n_3}}\chi^{(2)} = 2.411639\times10^{-3}, \left(\frac{\mu_0}{\varepsilon_0}\right)^{3/4} = \left(\frac{4\pi\times10^{-7}}{8.85\times10^{-12}}\right)^{3/4}$$

$$g_0 = 2\sqrt{\frac{\omega_3}{\omega_2}}|\kappa||A_1(0)| = 1.599701$$

For a single 1 cm step and $\Delta k = 0$:

$$C_2^{'} = \frac{1}{g}\left[-\Delta kA_2(0) + \sqrt{\frac{\omega_2}{\omega_3}}g_0A_3(0)\right] = \sqrt{\frac{\omega_2}{\omega_3}}A_3(0) = 0.$$

$$C_3^{'} = \frac{1}{g}\left[-\Delta kA_3(0) + \sqrt{\frac{\omega_3}{\omega_2}}g_0A_2(0)\right] = \sqrt{\frac{\omega_3}{\omega_2}}A_3(0) = 3.316625$$

$\cos(gz/2) = 0.6968141$, $\sin(gz/2) = 0.7172518$

$$A_2(z) = [A_2(0)\cos(gz/2) + iC_2^{'}\sin(gz/2)]e^{igz/2} = A_2(0)\cos(gz/2) + iC_2^{'}\sin(gz/2)$$
$$= 0.6968141$$

$$A_3(z) = [A_3(0)\cos(gz/2) + iC_3^{'}\sin(gz/2)]e^{-igz/2} = A_3(0)\cos(gz/2) + iC_3^{'}\sin(gz/2)$$
$$= 2.3836299i$$

$$\Delta A_1 = \frac{\omega_1}{\omega_2}\frac{1}{2g}\tilde{\kappa}^*\left\{i\left[A_2^*(0)A_3(0) - C_2'^* C_3'\right]\sin(g\Delta z) + \left[C_2'^* A_3(0) - C_3' A_2^*(0)\right]\left[1 - \cos(g\Delta z)\right]\right.$$
$$\left. + i\left[A_2^*(0)A_3(0) + C_2'^* C_3'\right]g\Delta z\right\}$$

$$\Delta A_1 = \frac{\omega_1}{\omega_2}\frac{1}{2g}\tilde{\kappa}^*\left\{-C_3' A_2^*(0)[1 - \cos(g\Delta z)]\right\} = -0.02572251A_1(0) + \Delta A_1 = 99.97427$$

$I_1 = |A_1|^2 = 9994.8407\text{w/cm}^2$, $I_2 = |A_2|^2 = 0.4855\text{w/cm}^2$, $I_3 = |A_3|^2 = 5.68169\text{w/cm}^2$.

9.8.2 Numerical Example, Calculation 2

$I_1 = |A_1|^2 = 10^4\text{w/cm}^2$, $I_2 = |A_2|^2 = 1\text{w/cm}^2$, $I_3 = |A_3|^2 = 0\text{w/cm}^2$, $\lambda_1 = 1.06\mu$, $\lambda_2 = 10.6\mu$, $\lambda_2 = 0.963636\mu$, $L = 1$ cm, $n_1 = n_2 = n_3 = 2.6$, $\chi^{(2)} = 1.1\times10^{-22}$ C/V^2.

From Eq. (9.231) and Eq. (9.232)

$$\frac{I_3(z)}{I_2(0)} = \frac{\lambda_2}{\lambda_3}\frac{\sin^2\left(\frac{g_0 z}{2}\right)}{\cos^2\left(\frac{g_0 z}{2}\right)\Big|_{z=0}} = \frac{\lambda_2}{\lambda_3}\sin^2\left(\frac{g_0 z}{2}\right). \tag{9.240}$$

From Eq. (9.220

$$g_0^2 = 2\frac{\omega_2\omega_1}{n_1 n_2 n_3}\left(\frac{\mu_0}{\varepsilon_0}\right)^{3/2}\left|\chi^{(2)}\right|^2|A_1|^2 = \frac{8\pi^2 c^2}{\lambda_3\lambda_2 n_1 n_2 n_3}\left(\frac{\mu_0}{\varepsilon_0}\right)^{3/2}\left|\chi^{(2)}\right|^2|A_1|^2 \tag{9.241}$$

$$g_0^2 = \frac{8\pi^2\times(2.99792458)^2\times10^{20}}{0.963636\times10^{-4}\times10.6\times10^{-4}\times(2.6)^3}\left(\frac{4\pi\times10^{-7}}{8.85\times10^{-12}}\right)^{3/2}\times\left(1.1\times10^{-22}\right)^2\times10^{-4} \tag{9.242}$$

$$g_0 = 0.01599701\text{ cm}^{-1},\ \sin\left(\frac{g_0 z}{2}\right) = 0.00799842\ . \tag{9.243}$$

From Eq. (9.240)

$$\text{Efficiency of conversion } \frac{I_3(z)}{I_2(0)} = \frac{10.6}{0.963636} 0.6397\times10^{-4} = 7.03722\times10^{-4} . \tag{9.244}$$

The conversion efficiency is in good, although not perfect, agreement with Yariv's result of 6×10^{-4} [10].

9.9 Birefringent Effects

The OPO treatment used here explicitly separates the kinetic and diffraction propagation effects. Continuing in this vein, we may include birefringent effects by considering only diffraction propagation and refraction. At this time only uniaxial crystals are considered. For a uniaxial crystal the index of refraction ellipse may be written:

$$\frac{x^2}{n_o^2} + \frac{y^2}{n_o^2} + \frac{z^2}{n_e^2} = 1 . \tag{9.245}$$

In terms of the angle of the crystal axis with respect to the propagation direction:

$$\frac{1}{n_e^2(\theta)} = \frac{\cos^2\theta}{n_o^2} + \frac{\sin^2\theta}{n_e^2} . \tag{9.246}$$

See Yariv, p88 [10].

$$\frac{dn_e(\theta)}{d\theta} = \frac{1}{2} n_e^3(\theta) \sin\left\{ 2\theta \left(\frac{1}{n_o^2} - \frac{1}{n_e^2} \right) \right\} . \tag{9.247}$$

See Shen, p128 [9]. The index of refraction in a small neighborhood of the extraordinary ray is[4]:

$$n_e(\theta_e + \Delta\theta) = n_e(\theta) + \left. \frac{dn_e(\theta)}{d\theta} \right|_{\theta_e} \Delta\theta . \tag{9.248}$$

The term $\Delta\theta$ is the incremental angle in the local angular neighborhood of the beam where the spread of k-vectors due to diffraction and aberration are significant. For a finite value of θ, an incremental angle $\Delta\theta$ may be written in terms of the paraxial angles α and β

$$\Delta\theta = \alpha \sin\psi + \beta \sin\psi , \tag{9.249}$$

where ψ is the azimuthal angle of the axis of the uniaxial crystal as it projects onto the α-β plane, similar to the x-y plane, and measured clockwise from the y-axis. Eq. (99.249) is not appropriate when the crystal axis is closely aligned to the optical axis.

Once ψ is defined, the polarization state may be decomposed into the e-ray which has the electric vector parallel to ψ and the o-ray which has the electric vector orthogonal to ψ. Since in GLAD we already have a

decomposition into x- and y-states, we may work with this x-y decomposition if the azimuthal angle of the crystal is constrained to be either 0º or 90º. For $\psi = 0^o$, x-polarization corresponds to the o-ray and y-polarization corresponds to e-ray. For $\psi = 90^o$, x-polarization corresponds to the e-ray and y-polarization corresponds to o-ray. The phase factor for diffraction propagation from Eq. (3.36) is

$$e^{jk_z z} = e^{jz\sqrt{k^2(1-\alpha^2-\beta^2)}} \approx e^{jkz} e^{-\frac{jkz}{2}(\alpha^2+\beta^2)} \quad . \tag{9.250}$$

The incremental angle $\Delta\theta$ is equivalent to α and β. Including the variation of index with angle from Eq. (9.248), but neglecting third order terms of the form

$$\frac{\partial}{\partial\alpha} n_e(\theta)\alpha^3 \text{ and } \frac{\partial}{\partial\beta} n_e(\theta)\beta^3 ,$$

we have

$$e^{jk_z z} = e^{j\frac{2}{\pi}\left[n_e(\theta) + \frac{dn_e(\theta)}{d\theta}(\alpha\sin\psi + \beta\cos\psi)\right]z} e^{-j\frac{2\pi}{\lambda}n_e(\theta)\frac{\alpha^2+\beta^2}{2}z} , \tag{9.251}$$

$$e^{jk_z z} = e^{jk_e z} e^{-jk_e\left[\frac{\alpha^2+\beta^2}{2} - \frac{1}{n_e(\theta)}\frac{dn_e(\theta)}{d\theta}(\alpha\sin\psi + \beta\cos\psi)\right]z} , \tag{9.252}$$

where $k_e = 2\pi n_e(\theta_e)/\lambda$. The linear terms in Eq. (9.252) causes the well-known walk-off effect. The birefringent walk-off causes the e-component of the beam to acquire a shear with respect to the e-ray direction due to diffraction propagation. The phase shift as a function of spatial frequency is

$$e^{j\frac{2\pi}{n_e(\theta)}\frac{dn_e(\theta)}{d\theta}(\xi\sin\psi + \eta\cos\psi)z} , \tag{9.253}$$

where ξ and η are the spatial frequency components. The walk-off as a function of propagation distance is

$$\text{walk-off} = \frac{1}{n_e(\theta)}\frac{dn_e(\theta)}{d\theta}(\hat{x}\sin\psi + \hat{y}\cos\psi)z . \tag{9.254}$$

In the case of KDP, $n_o = 1.51$ and $n_e = 1.47$ for $\theta = 20^o$ and $\frac{1}{n_e(\theta)}\frac{dn_e(\theta)}{d\theta} = -.0176$, giving a walk-off angle of about 17 milliradians.

If $\psi = 0$, the y-state of polarization is the e-ray and if $\psi = 90^o$ the x-state is the e-ray. It is necessary to set the index of refraction so n_o and n_e correspond to the x- or y-state of polarization appropriately. It is also necessary to include Eq. (9.246) for whichever state constitutes the e-ray.

9.9.1 Calculation Steps

The calculation proceeds via the split-step method in several parts:

- Conversion of relative phase to absolute phase by multiplying in $\exp[j\Delta\varphi]$ factor from OPL values.
- Point calculation of kinetic interactions with integrated k-vector mistuning by Eqs. (9.148)-(9.150) or Eqs. (9.206), (9.207), and (9.248).
- Propagation in birefringent media with different media wavelength for x-and y-polarization states and walk-off.

9.10 Second Harmonic Generation

Second harmonic generation satisfies the coupled differential equations[4],

$$\frac{\partial A_\omega}{\partial z} = i\tilde{\kappa}^* A_\omega^* A_{2\omega} e^{i\Delta kz}, \tag{9.255}$$

$$\frac{\partial A_{2\omega}}{\partial z} = i2\tilde{\kappa}^* A_\omega^2 e^{-i\Delta kz}, \tag{9.256}$$

where

$$\tilde{\kappa} = \frac{1}{\sqrt{2}}\omega\left(\frac{\mu_0}{\varepsilon_0}\right)^{3/4}\frac{1}{\sqrt{n_\omega^2 n_{2\omega}}}\chi^{(2)}. \tag{9.257}$$

This set of equations has been solved under general conditions by Armstrong[12]. We can prove that the total intensity ($I = 2I_\omega + I_\omega = 2|A_\omega|^2 + |A_{2\omega}|^2$) is a constant that is independent of z known as the Manley-Rowe Relation ($\partial I/\partial z = 0$). For now we assume the case of perfect phase matching so that $\Delta k = 0$. For a lossless medium $\tilde{\kappa}^* = \tilde{\kappa}$,

Let

$$u_1(z)e^{i\phi_1(z)} = A_\omega(z), \tag{9.258}$$

$$u_2(z)e^{i\phi_2(z)} = \frac{1}{2}A_{2\omega}(z). \tag{9.259}$$

Equations (9.255) and (9.256) now become

$$\frac{\partial u_1(z)}{\partial z} = -2\tilde{\kappa}u_1(z)u_2(z)\sin\theta(z), \tag{9.260}$$

$$\frac{\partial u_2(z)}{\partial z} = \tilde{\kappa}u_1^2(z)\sin\theta(z), \tag{9.261}$$

$$\frac{\partial\theta(z)}{\partial z} = \cos\theta(z)\left[\frac{u_1^2(z)}{u_2^2(z)} - 4u_2(z)\right], \tag{9.262}$$

where $\theta(z) = \phi_2(z) - 2\phi_1(z)$.

We can prove that

$$u_1^2(z) + u_2^2(z) = u_1^2(0) + u_2^2(0) = \text{constant} = \frac{I}{2}, \tag{9.263}$$

and from Eq. (9.262)

$$u_1^2(z)u_2(z)\cos(\phi_2(0) - 2\phi_1(0)) = \text{constant} = \Gamma, \tag{9.264}$$

$$\frac{du_2^2(z)}{dz} = \pm 2\tilde{\kappa}\sqrt{\left(\frac{I}{2} - u_2^2(z)\right)u_2^2(z) - \Gamma^2}, \tag{9.265}$$

$$\frac{du_1^2(z)}{dz} = -2\frac{du_2^2(z)}{dz} = \mp 4\tilde{\kappa}\sqrt{\left(\frac{I}{2} - u_1^2(z)\right)u_1^4(z) - \Gamma^2}. \tag{9.266}$$

Solutions of Eqs. (9.265) and (9.266) can be expressed in terms of the Jacobi elliptical functions. The solution becomes particularly simple for the special case in which the constant Γ is equal to zero. If $\Gamma = 0$ then from Eq. (9.281), $\theta(z) = 0$ and Eqs. (9.260) and (9.261)

$$\frac{\partial u_1(z)}{\partial z} = -2\tilde{\kappa}u_1(z)u_2(z), \tag{9.267}$$

$$\frac{\partial u_2(z)}{\partial z} = \tilde{\kappa}u_1^2(z). \tag{9.268}$$

With the solutions

$$u_1(z) = u_1(0)\,\mathrm{sech}[\sqrt{2}u_1(0)\kappa z], \tag{9.269}$$

$$u_2(z) = \frac{1}{\sqrt{2}}u_1(0)\tanh[\sqrt{2}u_1(0)\kappa z]. \tag{9.270}$$

We use Eqs. (9.269) and (9.270) to determine A_ω and $A_{2\omega}$. Moreover, if $A_\omega(0)$ is real, the solution takes the form

$$A_\omega(z) = A_\omega(0)\,\mathrm{sech}[\sqrt{2}A_\omega(0)\kappa z], \tag{9.271}$$

$$A_{2\omega}(z) = \frac{1}{\sqrt{2}} A_{\omega}(0)\tanh[\sqrt{2}A_{\omega}(0)\tilde{\kappa}z]. \tag{9.272}$$

The conversion efficiency in terms of photon flux

$$\frac{P_{2\omega}}{P_{\omega}} = \tanh^2[\sqrt{2}A_{\omega}(0)\tilde{\kappa}z]. \tag{9.273}$$

When $\tilde{\kappa}A_{\omega}(0)z \to \infty$, $P_{2\omega}/P_{\omega} \to 1$ and all the input photons can be converted into half as many output photons (at twice the frequency).

The penalty for deviating from the index-matching condition at a fixed L is a reduction of the second harmonic power output by the factor

$$\frac{P_{2\omega}}{P_{\omega}} = \frac{\sin^2\left(\frac{\Delta kL}{2}\right)}{\left(\frac{\Delta kL}{2}\right)^2}. \tag{9.274}$$

We write the differential equations in terms of photon flux amplitude $|p|^2 = P$. Second harmonic generation satisfies the coupled differential equations[4],

$$\frac{dp_2}{dz} = -j\frac{\kappa}{2\sqrt{2}} p_1^2 e^{j\Delta kz}, \tag{9.275}$$

$$\frac{dp_1}{dz} = -j\frac{\kappa}{\sqrt{2}} p_2 p_1^* e^{-j\Delta kz}, \tag{9.276}$$

where p_1 and p_2 are the photon flux complex amplitudes of the pump and second harmonic beams, such that $|p|^2$ has units of photons/cm2 and

$$\kappa = 2\tilde{\kappa}\sqrt{\frac{hc}{\lambda_{\omega}}} = 2\tilde{\kappa}\sqrt{\hbar\omega}. \tag{9.277}$$

The energy per photon is

$$\text{energy per photon} = \frac{hc}{\lambda} = \frac{19.8762\times10^{-24}[\text{J/cm}]}{\lambda[\text{cm}]}, \tag{9.278}$$

$$p_1 = A_1\sqrt{\frac{\lambda_1}{hc}},\ p_2 = A_2\sqrt{\frac{\lambda_2}{hc}}. \tag{9.279}$$

The coupling constant has the units $\text{cm}\sqrt{\text{photon}}$. The detuning factor reduces the coupling coefficient. When the frequency doubling is broken up into several intervals using separate calls to `double`, the cumulative effect of detuning may be accomplished by setting Zstart to the length of previous passes through the doubler. The gain length for the power conversion is

$$\text{gain length} = \frac{1}{\kappa}\sqrt{\frac{hc}{\lambda I_1(0)}}, \tag{9.280}$$

where the wavelength is expressed in centimeters and I_1 is the incident intensity in watts/cm2 of the pump. For I_1 = 10 Mw/cm2, $\lambda_1 = 1.06\times10^{-4}$ cm, and $\kappa = 5\times10^{-4}$, the gain length is 1.83 cm. For $A_2(0) = 0$ and no detuning, the solution is

$$I_2(z) = \frac{I_1(0)}{2}\tanh^2\left(\frac{\kappa}{2}\sqrt{I_1(0)\frac{\lambda_1}{hc}}z\right). \tag{9.281}$$

Detuning may be included approximately in the above expression by

$$\kappa \to \kappa\frac{\sin^2\left(\frac{1}{2}\Delta kz\right)}{\left(\frac{1}{2}\Delta kz\right)^2}. \tag{9.282}$$

9.11 Transient Raman Kinetics

Very high power, short pulse laser beams are important for certain types of laser fusion. The Omega laser being developed by the Laboratory for Laser Energetics, University of Rochester employs many beams at power levels of approximately 2 gigawatt per square cm (as of 1992, Omega upgrade). Propagation of these beam through air may induce rotational Raman scattering to form a Stokes beam, which given sufficient distance for amplification, may significantly deplete the initial beam. For high power beams, the Stokes beam may grow quite quickly and, as a consequence, the beam quality of the Stokes may be extremely poor, making it essentially useless.

To facilitate design and analysis of these high power systems, it is very helpful to have an accurate computer model so that different configurations and laser inputs can be easily studied. We must resort to a computer model rather than rely entirely on analytical solutions because analytical descriptions in the literature primarily address relatively weak amplification such that the Raman conversion takes place a distances comparable to a Fresnel number of 1 or less.

A single leg of the Omega laser has a diameter of approximately 30 cm and may see significant Raman conversion in over the range of between 20 and 40 meters. Some typical parameters are listed in Table 9.1. At 0.351μ the Fresnel number is on the order of 2,000 or more.]

To model this system, we have several difficult issues of physics:

- transient Raman equations rather than the simpler, steady-state equations,
- correct treatment of noise to generate the Stokes beam,

Table. 9.1. Typical parameters for a high power laser exhibiting Raman conversion effects .

Parameter	Value
k_1	4.5647 cm sec gm^{-1}
k_2	0.1443 cm^{-1}
Γ	7.52×10^{0}
N	2.6856×10^{19} cm^{-3}
k_3	0.1447
g_1	6.76 cm/terrawatt
λ_L	3.511×10^{-5} cm
λ	3.5204×10^{-5} cm
energy per pulse	500 J
beam diameter	30 cm
pulse length	100 to 500 picoseconds

- representation of the wide angle radiation typical of Stokes light in high gain, short distance configurations.

We wish to have our model be efficient which means we must pay careful attention to the size of the arrays required and the number of axial steps. The basic equations for the evolution of the radiation and the state of the medium (Raymer and L. A. Westling, [5]) are

$$\left[\nabla_\perp^2 + 2ik_l\frac{\partial}{\partial z}\right]E_l = 2k_3k_lQE_s\,, \tag{9.283}$$

$$\left[\nabla_\perp^2 + 2ik_s\frac{\partial}{\partial z}\right]E_s = 2k_3k_sQ^*E_l\,, \tag{9.284}$$

$$\frac{\partial Q^*}{\partial t} = -\Gamma Q^* + ik_1E_l^*E_s + F^*\,, \tag{9.285}$$

where E_l and E_s are the complex amplitude fields of the laser pump and Stokes radiations; Q is the complex amplitude state of the medium excitation, k_l and k_s are the wave numbers; Γ is the medium linewidth; k_1, k_2, and k_3 are coupling coefficients; and F is the random force due to quantum effects. The coefficients k_2 and k_3 are related by,

$$k_3 = \frac{\lambda_s}{\lambda_l}k_2\,. \tag{9.286}$$

Split step methods are invoked to solve these equations. In the split step method, we break the coupled equations into a series of elementary linear differential equations and solve the full set for each axial step, See Section 3.4. The effects of diffraction are included by solving,

$$\frac{\partial E_l}{\partial z} = \frac{i}{2k_l}\nabla_\perp^2 E_l \qquad \text{diffraction for } E_l, \tag{9.287}$$

$$\frac{\partial E_s}{\partial z} = \frac{i}{2k_l}\nabla_\perp^2 E_s \qquad \text{diffraction for } E_s. \tag{9.288}$$

We shall solve these diffraction equations with FFT techniques by preference. In our application of particular interest, diffraction propagation does not strongly effect E_l but does influence the Stokes beam because of its high spatial frequency components.

The kinetics equations to be solved are

$$\frac{\partial}{\partial z}E_l = ik_3 Q E_s, \tag{9.289}$$

$$\frac{\partial}{\partial z}E_s = -ik_2 Q^* E_l, \tag{9.290}$$

$$\frac{\partial Q^*}{\partial t} = -\Gamma Q^* + ik_1 E_l^* E_s + F^*, \tag{9.291}$$

where E_l and E_s are the laser and Stokes complex amplitudes, Q is the medium complex state, the k's are gain coefficients, Γ is the Raman line width, and F s a random force due to collisions.

Let us now consider the solution of Eq. (9.291). The second term on the right hand side represents the finite time response of the medium. The first term represents stimulated Raman gain. This term is driven by the product of the two optical fields and is not present in the absence of the driving laser pump. The third term on the right is due to collisional effects in the air and this force is always present irrespective of the presence of the laser field.

The collisional force is delta correlated in space and time and this leads to scattering of the laser field into a Stokes field which radiates into 4π steradians. F is delta-correlated in time and space

$$\langle F^*(\boldsymbol{r}, \tau), F(\boldsymbol{r}', \tau')\rangle \propto 2\frac{\Gamma}{N}\delta^3(\boldsymbol{r} - \boldsymbol{r}')\delta(\tau - \tau'). \tag{9.292}$$

Let us consider that Q consists of two components: a stochastic component due to F and a stimulated Raman component due to E_l and E_s,

$$Q = Q_{\text{raman}} + Q_{\text{ss}}, \tag{9.293}$$

where Q_{raman} is driven by stimulated Raman and Q_{ss} is the stochastic component induced by collisional forces.

Since F is a statistically stationary, Q_{ss} is statistically stationary. However, unlike F, Q varies with finite temporal correlation as determined by Γ.

To understand the effect of Q_{ss}, let us first completely neglect the effect of stimulated Raman. Consider Fig. 9.6, where Q_{ss} is shown as a random field of complex numbers. The random field is delta correlated spatially but varies temporally at the rate Γ. If a weak laser field (one weak enough to have negligible Raman amplification) passes through the field of scattering sites, spontaneous Raman will be generated. Because of the delta correlation of Q, the Stokes light will scatter in all directions. However, only the Stokes radiation that travels approximately in the forward direction will be amplified. Considering only the forward scattered light in some small solid angle $\Delta\omega$, the Stokes noise power will increase linearly due to spontaneous Raman. This takes the form

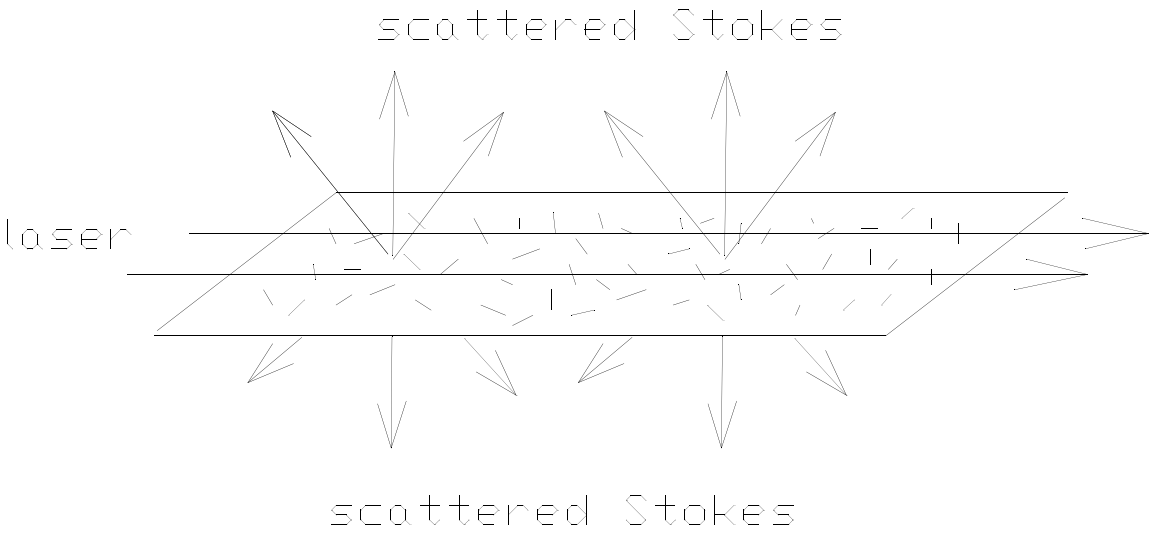


Fig. 9.6. A laser beam incident from the left is scattered by the medium field Q_{ss} which is delta correlated spatially and which scatters light into steradians. In the scattering due to spontaneous Raman effects obeys Beer's Law. Only the light which is scattered in the forward direction contributes to the amplified Stokes.

$$\langle I_s(x, y, z + \Delta z, t_n)\rangle \;=\; \langle I_s(x, y, z, t_n)\rangle + g_2 I_l(x, y, z, t_n)\frac{\Delta\Omega}{4\pi}\Delta z\,, \tag{9.294}$$

where g_2 is a coefficient representing the spontaneous emission growth. We may construct a contribution to the Stokes complex amplitude E_s which represents the appropriate noise accumulated in a step of length Δz

$$E_s(x, y, z + \Delta z, t_n) \;=\; E_s(x, y, z, t_n) + \left[g_2 I_l(x, y, z, t_n)\frac{\Delta\Omega}{4\pi}\Delta z\right]^{1/2} X(x, y, z, t_n)\,, \tag{9.295}$$

where $X(x, y, z, t_n)$ is a unit variance random phasor with normally distributed magnitude. The stochastic variable $X(x, y, z, t_n)$ is readily constructed to be unit variance and spatially delta correlated. Since the field $X(x, y, z, t_n)$ represents the stochastic aspect of Q, it changes such that the temporal correlation is $1/\Gamma$. To represent this temporal behavior we shall add a small random component at each step in time. This takes the form,

$$X(x, y, z, t_n) = X(x, y, z, t_{n-1})e^{-\Gamma\Delta t} + Y(x, y, z, t_n)(1 - e^{-2\Gamma\Delta t})^{1/2}, \tag{9.296}$$

where $Y(x, y, z, t_n)$ is a new unit-variance, normally distributed random phasor. Eq. (9.296) gives a random field which evolves in time at the appropriate rate.

E_s is also driven by the stimulated Raman effect. In solving for the effect of the stimulated Raman, we neglect the spontaneous Raman effects. E_s is driven by the time-integrated effect of the laser field. This takes the form,

$$\frac{\partial}{\partial z}E_s = -ik_2 Q^* E_l \qquad \text{amplification of Stokes,} \tag{9.297a}$$

$$Q^*(x, y, z, t + \Delta t) = ik_1 E_l^*(x, y, z, t)E_s(x, y, z, t)\Delta t + e^{-\Gamma\Delta t}Q^*(x, y, z, t) \qquad \text{evolving sum,} \tag{9.297b}$$

$$\Delta H(z + \Delta z) = |E(z + \Delta z)|^2 - |E(z)|^2 \qquad \text{energy gain for Stokes,} \tag{9.297c}$$

$$E_l(z + \Delta z, t_n) = E_l(z, t_n)\sqrt{1 - \frac{\lambda_s}{\lambda_l}\frac{\Delta H}{|E_l(z)|^2}} \qquad \text{energy loss from pump.} \tag{9.297d}$$

We may also define $g_1 = 2k_1k_2/\Gamma$ as the steady-state power gain coefficient to be found when considering Eq. (9.297a) and Eq. (9.297b). In order to minimize the number of axial samples, GLAD takes advantage of the nearly (but not exactly) exponential behavior of Q, as shown in Fig. 9.7. To achieve pump depletion, we simple calculate the energy loss due to spontaneous emission (into 4π steradians, not just $\Delta\Omega$) and due to stimulated Raman amplification. GLAD employs photon conserving algorithms in the region of strong pump depletion.

We have left up to now the question of the appropriate value of for $\Delta\Omega$. For a uniform intensity pump beam, this value depends on the ratio of the diameter of the laser beam to the length. Consider Fig. 9.8, we have shown an initial plane, a propagation distance of E_l, and a final plane of observations. The path length has been divided into gain-lengths. For a beam having gigawatts of power per square centimeter, we have one or two gain lengths per meter of length. We assume that the most of the spontaneous Stokes growth occurs at the beginning of the propagation in the first gain-length. Of course spontaneous Stokes growth occurs at all axial regions, but the upstream noise experiences the greatest Raman amplification and is, therefore, the most important. The limits of the aperture define the maximum angles which can contribute to the Stokes beam at the end of the propagation. Alternately, we could consider the angles as they originate from the start of the propagation. Rays which leave the beam before the last one or two gain lengths do not receive enough amplification to remove much power from the laser beam, as indicated in Fig. 9.9. In practice

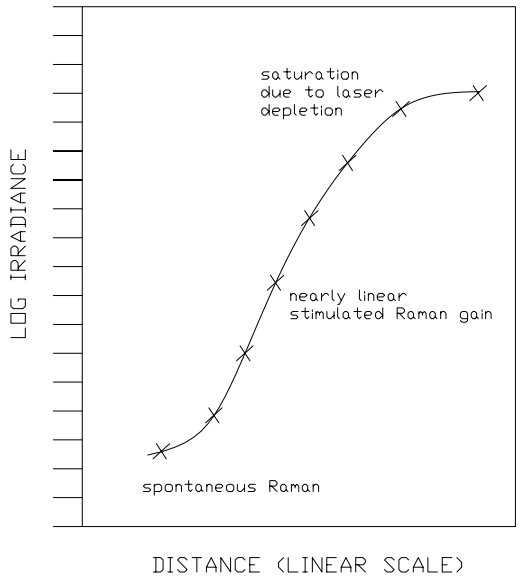


Fig. 9.7. Semi-exponential growth of Stokes light in the spontaneous regime. Shown schematically.

we should include angles somewhat larger than the diameter-to-length ratio to get the correct pump depletion and the Stokes power scattered well outside the diameter of the laser beam.

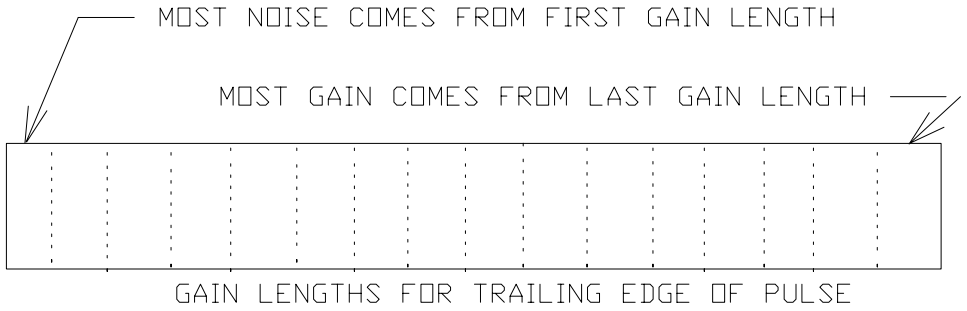


Fig. 9.8. The propagation length may be considered to be divided into gain-lengths. A strong beam will have many gain lengths. While spontaneous Raman noise is generated at all points in the beam, the noise from the first gain length has the highest stimulated gain and is, therefore, most important. Most of the power gain, however, occurs in the final gain length.

The diameter-to-length ratio gives a good value for the maximum angle only if the laser beam is uniform across the aperture. If there are hot spots in the laser beam or even relatively weak intensity peaks, very strong hot spots will be induced into the Stokes beam because of the very strong amplification. Since one hot spot will generally dominate, the effective width of the Stokes beam may be much less than the full aperture. This will result in considerable narrowing of the angles which must be modeled and, correspondingly, a substantial increase in speckle size.

To better understand the concept of the effective source, we can estimate its value. For collimated geometry and weak pump depletion, we have the noise growing as

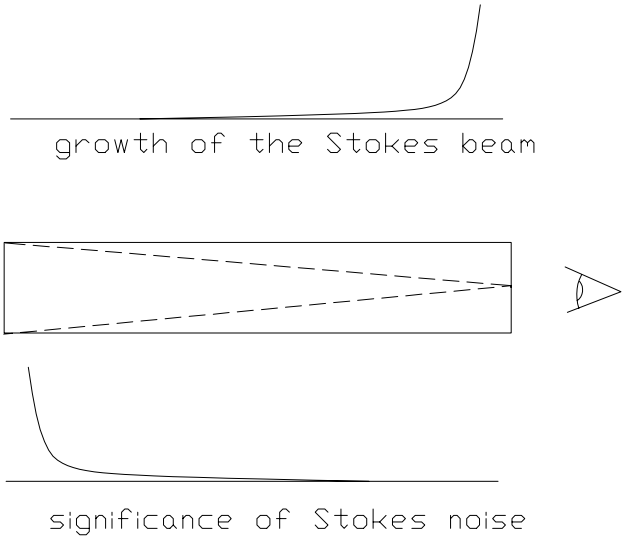


Fig. 9.9. The Stokes noise is amplified by the stimulated Raman gain, shown as an exponentially increasing function in the upper figure. If we were to look into the beam path (neglecting safety considerations and the fact that we can not see UV light), we would see the most distant noise as being the brightest and noise generated closer to the eye as being of negligible importance.

$$\frac{\partial I_s}{\partial z} = g_2 \frac{\Delta\Omega}{4\pi} I_l, \tag{9.298}$$

and the stimulated amplification as

$$\frac{\partial I_s}{\partial z} = g_t I_l I_s, \tag{9.299}$$

where g_t is the transient stimulated Raman gain, which depends on the laser temporal waveform and coherence. Ignoring diffraction effects, the solution is

$$I_s(L) = \int_0^L g_2 \frac{\Delta\Omega}{4\pi} I_l e^{g_t I_l (L-z)} dz, \tag{9.300}$$

which has the solution,

$$I_s(L) = \frac{g_2}{g_t} \frac{\Delta\Omega}{4\pi} (e^{g_t I_l L} - 1). \tag{9.301}$$

This equation shows that the Stokes irradiance is closely approximated by $\frac{g_2}{g_t}\frac{\Delta\Omega}{4\pi}$. We can always consider the effective starting noise to be

$$I_s(0) = \frac{g_2}{g_t}\frac{\Delta\Omega}{4\pi}. \tag{9.302}$$

Or more generally,

$$I_s(0) = \frac{g_2}{g_t}\frac{A}{4\pi L^2}, \tag{9.303}$$

where A is the area of the beam. It is interesting to note that the effective starting noise is independent of the laser pump irradiance. The only variable is the geometrical factor $A/4\pi L^2$, since both g_t and g_2 are fundamental properties of the atmosphere (actually g_t also varies with the laser temporal waveform and coherence).

The Raman threshold (by the usual conventions) is

$$I_s(L) = 0.01 I_l, \tag{9.304}$$

from which we conclude that the exponent of the amplification at threshold is

$$\ln\left[\frac{0.01 g_t I_l}{g_2}\frac{4\pi}{\Delta\Omega} + 1\right] = g_t I_l L \quad \text{(exponent of amplification at threshold).} \tag{9.305}$$

Clearly the amplification necessary to achieve threshold in a fixed geometry varies with laser pump intensity. If we include the length of the beam path as a variable we have,

$$\ln\left[\frac{0.01 g_t I_l}{g_2}\frac{4\pi L}{A} + 1\right] = g_t I_l L. \tag{9.306}$$

Interestingly, the solution to this transcendental equation gives a value for the amplification exponent at threshold which depends only on the geometry and the spontaneous Raman gain coefficient g_2.

From these calculations we can readily see that there is an effective source which due to the spontaneous emission and the stimulated amplification. Note that slight irregularities in E_l will cause large variations in amplification because they occur in the exponent. The amplified effective source will appear to be much smaller than the diameter of the beam given a single weak hot spot in the laser which is brighter than all other spots.

Since E_s scatters in all directions, the usual split-step method is called into question, because the diffraction propagation methods are suitable only for near-diffraction-limited beams. With a numerical array and sample spacing of Δx and Δy, the maximum solid angle which can be represented is $\Delta\Omega = \lambda^2/\Delta x \Delta y$. For a large diameter beam and short propagation length, the angle represented by the diameter-to-length ratio may be many times larger than the angle which can be represented by an array of reasonable size, as

illustrated in Fig. 9.10. In the more usual case, where hot spots dominate, the speckle size is much larger, as indicated in Fig. 9.11, and may be resolved more easily.

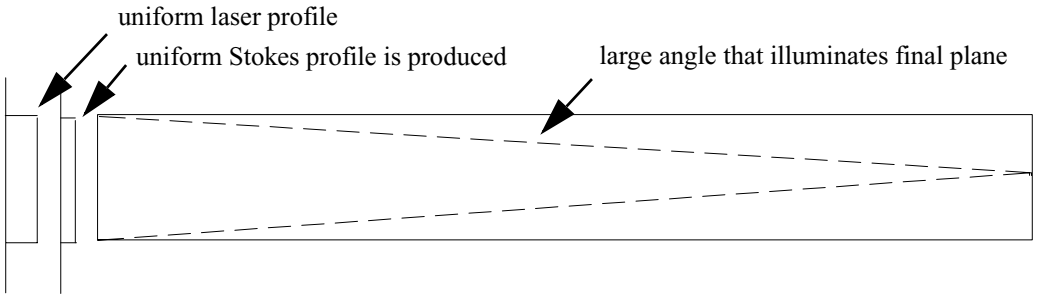


Fig. 9.10. A perfectly uniform laser beam will generate Stokes noise which has a uniform envelope. This represents a coherent delta correlated source which illuminates the plane of observation. The angle subtended by the source determines the amount of noise contributing to the process and also determines the speckle size. The effective source is approximately the noise generated in the first gain length times the amplification.

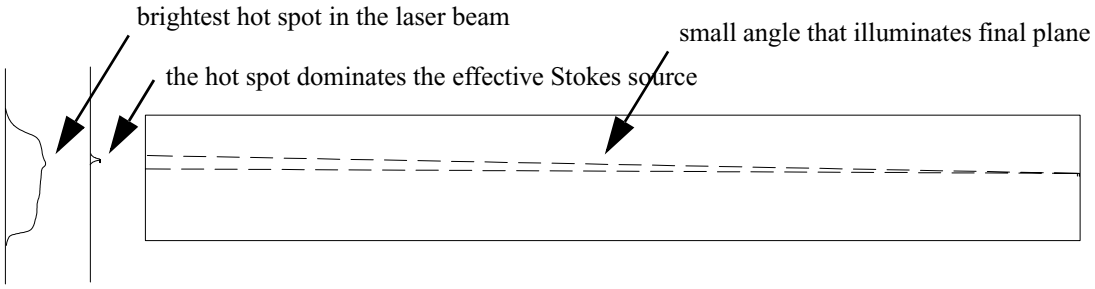


Fig. 9.11. .Because of the very high gain, slight variations in the laser intensity become greatly exaggerated in the Stokes beam. The point of highest intensity in the laser creates a single hot spot in the amplified Stokes beam which may be much smaller than the diameter of the beam leading to a greatly reduced speckle size. Because of the much greater speckle size, the sampling required is easier to achieve.

We shall limit the angular subtense of E_s by deleting the high spatial frequency components of F to introduce only angles of E_s which fit within the solid angle $\Delta\Omega$.

References

1. Briggs, Warner, Lawrence, Barnard, and Bobbs, “Raman Resonator Modeling and Analysis Basic Program Final Report,” Rocketdyne Division, Rockwell International.
2. “Scaling Studies of Efficient Raman Converters, Final Report,” July 1, 1983, NRTC-83-09R, Northrop.
3. Gruhl and Sigel, IEEE Journal of Quantum Electronics, Vol 20, No. 9, September 1984.
4. A. Yariv, *Introduction to Optical Electronics*, Holt, Rinehart, and Winston, New York, (1976).
5. M. G. Raymer and L. A. Westling, “Quantum theory of Stokes generation with a multimode laser,” JOSA B, Vol. 2, No. 9, p1417-1421 (1985).
6. T. H. Maiman, “Stimulated optical radiation in ruby masers,” Nature, (London) **187**, 493 (1960).

7. Y-K Kuo, H-M Chen, J-Y Chang, “Numerical study of the Cr:YSO Q-switched in ruby laser,” Opt. Eng. **40**(9) 2031-2035 (Sept. 2001).
8. Walter Koechner, *Solid-State Laser Engineering*, Springer-Verlag (1996).A. V. Smith, W. J. Alfred, and T. D. Raymond, “Comparison of a numerical model with measured performance of seeded nanosecond KTP optical parametric oscillator”, JOSA B, Vol. 12, No. 11, 2253–2267 (1995).
9. Y. R. Shen, *The Principles of Nonlinear Optics*, John Wiley & Sons, New York (1984).
10. A. Yariv, *Quantum Electronics*, 2nd Ed., John Wiley & Sons, (1975).
11. J. A. Fleck, Jr. and M. D. Feit, “Beam propagation in uniaxial anisotropic media”, JOSA, Vol. 73, No. 7, 920–926, July 1983.
12. J. A. Armstrong, N. Bloemberger, J. Ducuing, and P. S. Pershan, Phys. Rev., Vol. 127, 1918 (1962).

10. Component Specification

GLAD allows modeling of diverse forms of mirrors, lenses, and apertures.

10.1 Mirrors and Lenses

GLAD models a mirror by its radius (or focal length) using the `mirror` command. If the `mirror/global` command is used the component will be modeled using exact rays to calculate OPD values and global positioning, as discussed in Chap. 11. The quadratic phase is

$$t(x, y) = e^{-j\frac{k}{2f}(x^2 + y^2)}, \tag{10.1}$$

where $t(x, y)$ is the complex transmission function, k = the wave number, f is the mirror radius.

The sign convention for focal length and radius are different. The focal length is defined with respect to the element's effect on the beam:

$f > 0$ causes the light to be more convergent,

$f < 0$ causes the light to be more divergent.

The sign convention of the mirror radius is defined by the position of the center of curvature with respect to the mirror vertex, in the vertex coordinate system. Convex mirrors have positive radii. Concave mirrors have negative radii. The global coordinate command, `vertex/rotate`, rotates the vertex coordinate system with respect to the global system. It is possible to rotate a concave mirror so that the beam hits the backside, causing the element to have the effect of a convex mirror. The user should exercise care in laying out elements with the `vertex/rotate` command to be sure the correct sign conventions are used. Remembering that these conventions are defined in the vertex coordinate system:

$r > 0$ center of curvature to right of vertex,

$r < 0$ center of curvature to left of vertex.

Idealized lenses are defined similarly to paraxial mirrors, using the `lens` command. Detailed ray tracing, discussed in Chap. 11, is treated with the `lensgroup` command. The optical power of the lens is defined by the focal length.

For either lenses or mirrors the phase imposed by the element may be applied as a change to the complex amplitude distribution or as a change to the phase bias. The choice is determined by GLAD to achieve the minimum phase curvature that exists in the complex amplitude to minimize sampling requirements.

Lenses and mirrors may be defined with spherical, cylindrical, or toroidal elements. In the paraxial approximation, these elements are treated only to second order, so no spherical aberration is imposed. Phase bias is represented to second order with an R_x and R_y. The command `abr/focus` may be used to add phase curvature to the complex amplitude distribution without changing the phase bias. `abr/sph` may be used to add spherical aberration.

As mentioned above, the `mirror/global` command models conic mirrors using exact OPD calculations. The phase bias is treated as an exact toric phase surface with radii, R_x and R_y.

10.2 Tilts and Decenters

Tilts and decenters are divided to fine scale and large scale. Fine scale tilts and decenters may be treated by the `abr/tilt` and either the `shift` or `rescale/shift` commands. These commands may be used for slight misalignments of an optical system or resonator. `abr/tilt` directly changes the phase of the distribution in the complex amplitude distribution represented in the computer array. `shift` or `rescale/shift` implements translations directly in the complex amplitude array. Large scale rotations and translations used for defining a folded configuration may be defined with the `global` and `vertex` commands in the next chapter.

10.3 Apertures and Obscuration

Apertures and obscurations may be modeled with the `clap` and `obs` commands. These elements are applied normal to the propagation direction of the beam. A general obscuration may be applied with the `obs/gen` command where the coordinates of the vertices of the obscuration are read from a file.

10.4 Complex Systems and GRIN Media

Propagation through a complex system or through a gradient refractive index (GRIN) medium can be represented by an ABCD matrix. An equivalent, elementary optical system can be found for any ABCD matrix. Let the properties of the first optical space be indicated by the subscript 1 and the properties of the second space by the subscript 2. It is assumed that the ABCD matrix consists of four elementary operations: change of index, change of magnification, thin lens of optical power ϕ, and a translation of distance t:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M & 0 \\ 0 & 1/M \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\phi/n_2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & n_1/n_2 \end{bmatrix}. \tag{10.2}$$

One could choose a different set of four primitive operations, for example, by permuting the order of the matrices, and develop the method in similar fashion to that done here. Siegman describes a generalized diffraction operator using these ABCD matrices explicitly — an alternate approach. The elementary operator method is preferred because of the physical insight to be derived and because the simple operations already exist in most propagation codes. The elementary operations are illustrated in Fig. 10.1.

Equation (10.2) may be interpreted as finding the image of the original distribution in the local optical space, and completing the propagation to the desired plane. There are no diffraction propagation effects in transforming the object space distribution to the local image —only magnification and radius change. All diffraction propagation effects will occur in the single propagation step. The solution for the four free variables is

$$M = \frac{AD - BC}{D}, \tag{10.3}$$

$$\frac{n_1}{n_2} = AD - BC, \tag{10.4}$$

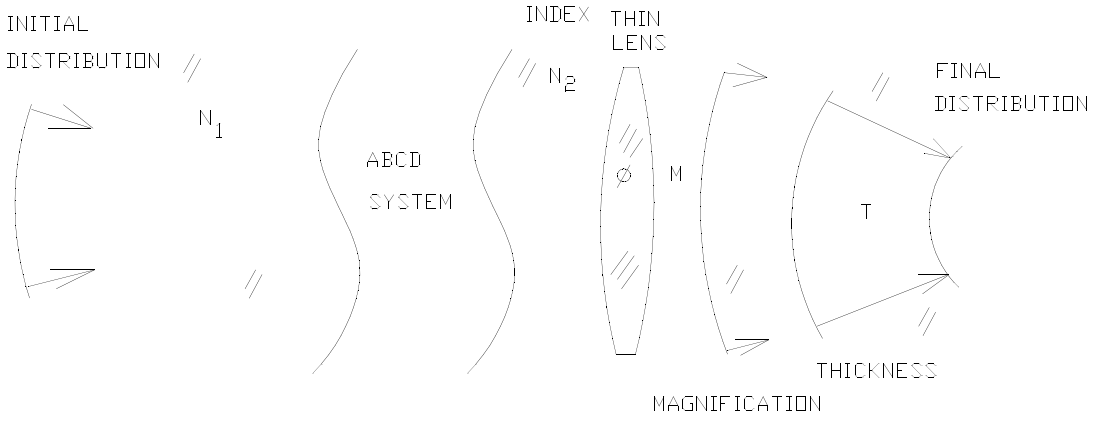


Fig. 10.1. Propagation through a complex paraxial optical system can be represented by an equivalent elementary optical system.

$$\frac{\phi}{n_2} = -\frac{CD}{AD - BC} = -MC \quad , \tag{10.5}$$

$$t = \frac{B}{D}. \tag{10.6}$$

The first three steps are quite simple to execute. In numerical implementation, a magnification change alters the size of the sample spacing, $\Delta x \rightarrow M\Delta x$, and the reference radius (if one is used), $R_1 \rightarrow M^2 R_1$. The refractive index change does not change the wavefront error of the complex amplitude distribution but does change the reference radius to $\frac{n_2}{n_1} M^2 R_1$. The thin lens changes the radius of curvature to

$$\frac{1}{R'_1} = \frac{1}{\frac{n_2}{n_1} M^2 R_1} - \phi. \tag{10.7}$$

The last step is a propagation of t in general, noncollimated geometry. Only this step requires any substantial computation time. In general, noncollimated propagation, the beam may be referenced either to a reference plane or a reference sphere þ which ever minimizes the quadratic phase error in the complex amplitude distribution.

10.4.1 Alternate form

If the element D is too small — due to the original distribution being projected to infinity in image space — an alternate form provides better numerical accuracy, as shown in Fig. 10.2.

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\phi_1/n_2 & 1 \end{bmatrix} \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\phi_2/n_2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & n_1/n_2 \end{bmatrix}, \tag{10.8}$$

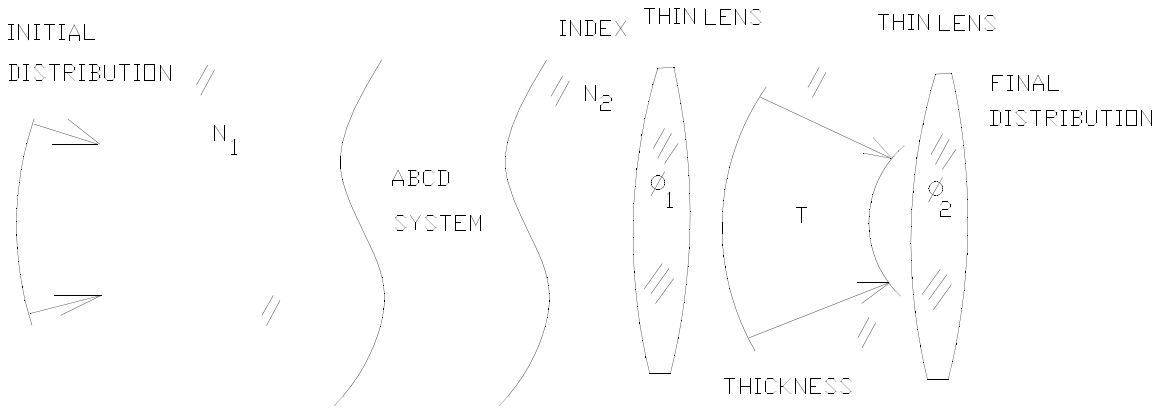


Fig. 10.2. Alternate form of equivalent elementary optical system more suited to having the image of the distribution projected to infinity.

$$\frac{n_1}{n_2} = AD - BC, \tag{10.9}$$

$$t = \frac{B}{AD - BC}, \tag{10.10}$$

$$\frac{\phi_1}{n_2} = \frac{1 - A}{t}, \tag{10.11}$$

$$\frac{\phi_2}{n_2} = \frac{\dfrac{\phi_1}{n_2} - C}{A}. \tag{10.12}$$

10.5 Anamorphic ABCD matrices

The optical system may not be rotationally symmetric. If there is bilateral symmetry, the paraxial behavior may be described by two matrices for the two orientations,

$$\begin{bmatrix} A_x & B_x \\ C_x & D_x \end{bmatrix} = \begin{bmatrix} 1 & t_x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M_x & 0 \\ 0 & 1/M_x \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\phi_x/n_2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & n_1/n_2 \end{bmatrix}, \tag{10.13}$$

$$\begin{bmatrix} A_y & B_y \\ C_y & D_y \end{bmatrix} = \begin{bmatrix} 1 & t_y \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M_y & 0 \\ 0 & 1/M_y \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\phi_y/n_2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & n_1/n_2 \end{bmatrix}. \tag{10.14}$$

We require isotropic media and therefore constrain the refractive indices to be the same in the two directions. It is desirable to put the equations for the two directions in a form which has the same optical propagation thickness. We can propagate t_x for both directions by pre- and post-multiplying the y-matrix for distance propagation

$$\begin{bmatrix} 1/M_2 & 0 \\ 0 & M_2 \end{bmatrix} \begin{bmatrix} 1 & t_x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M_2 & 0 \\ 0 & 1/M_2 \end{bmatrix} = \begin{bmatrix} 1 & t_y \\ 0 & 1 \end{bmatrix} \text{ where } M_2 = \sqrt{\frac{t_x}{t_y}}, \tag{10.15}$$

such that the full y-matrix may be written in terms of t_x

$$\begin{bmatrix} A_y & B_y \\ C_y & D_y \end{bmatrix} = \begin{bmatrix} 1/M_2 & 0 \\ 0 & M_2 \end{bmatrix} \begin{bmatrix} 1 & t_x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M_2 & 0 \\ 0 & 1/M_2 \end{bmatrix} \begin{bmatrix} M_y & 0 \\ 0 & 1/M_y \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\phi_y/n_2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & n_1/n_2 \end{bmatrix}. \tag{10.16}$$

This allows propagation of the same distance for both x- and y-directions by pre-and postmultiplying by a magnification operator.

11. Global Coordinate System

This chapter describes modeling components in GLAD using the global coordinate system. The basis of GLAD, like other physical optics codes, was a quasi-paraxial approach which treated folded systems by implicitly unfolding the system and which could treat only small tilts and decenters. The global coordinate system is a much more powerful approach which allows components to be located at arbitrary positions in space and which allows large angle rotations. When fully implemented, the global coordinate system will allow calculations in three-dimensional space with correct treatment of optical aberrations of tilted surfaces and with accurate calculations of physical and optical path differences. Advantages of the global coordinate system and associated functions are:

- Global positioning of components,
- Arbitrary component rotations,
- Accurate path length calculations,
- Correct calculation of beam azimuthal rotations,
- Exact or approximate aberrations of tilted components,
- Surface polarization effects in s- and p-directions.

Systems which require this type of sophistication in the code include:

- Free-electron lasers (total path length is important),
- Systems with strongly tilted components,
- Coupled resonators,
- Complex folded resonators,
- Polarization calculations.

The `mirror/global`, `surface`, `lensgroup`, `clap`, and `obs` commands use the global coordinate system. The idealized lens and mirror commands are positioned on the chief ray of each beam.

11.1 Component Location and Rotation in Global Coordinates

To describe the propagation of a beam through a complex three-dimensional optical system, we define four coordinate systems

- global ray and vertex locations,
- ray complex amplitude distribution,
- vertex component rotations and shape definition,
- surface surface at chief ray intercept point.

When the optical elements are to be located in arbitrary positions with arbitrary rotations, it is necessary to define a global coordinate system on which the beam path and the optical components are specified. As the beams propagate through the optical system, the global coordinate system is updated. In GLAD, the location of the optical component is defined by the vertex location. The rotation is considered to be about the vertex location. The direction of the chief ray must also be defined on the same global coordinate system. It is also necessary to determine the azimuthal orientation of the complex amplitude distribution about the

chief ray. Systems with out-of-plane components can rotate the beam distribution about the chief ray direction. The azimuthal rotation may be determined by establishing a ray matrix consisting of i-, j-, and k-vectors. The k-vector defines the chief ray direction.

The properties of the optical component are most easily defined in terms of a vertex coordinate system. For a rotationally symmetric surface, the z-axis is identical to the axis of symmetry. A fourth coordinate system at the point of intersection of a ray allows calculation of the polarization properties. This surface coordinate system consists of s-, p-, and n-vectors. The n-vector is the surface normal vector. We choose the surface normal vector to point toward the center of curvature. Figure 11.1 illustrates the four coordinate systems schematically.

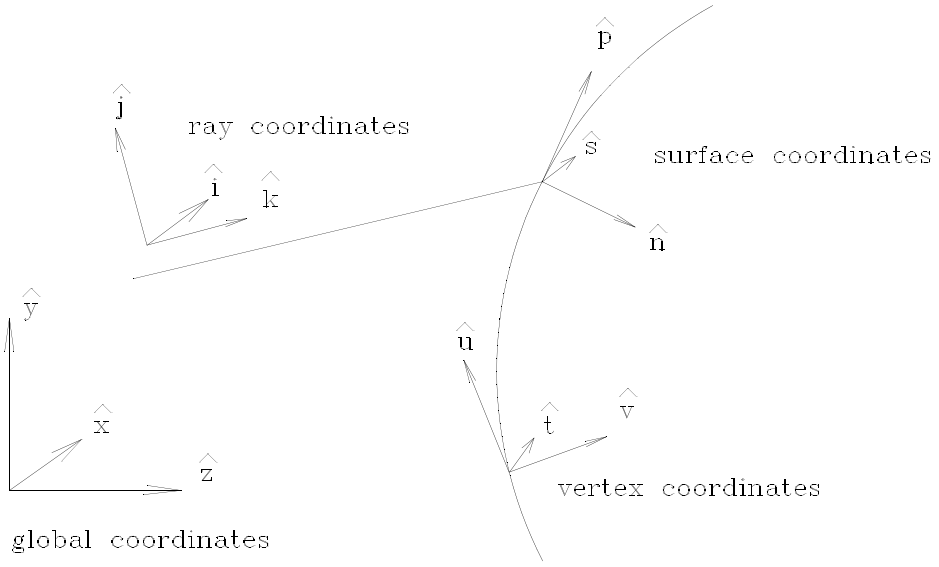


Fig. 11.1. The global, ray, vertex, and surface coordinate systems.

The unit vectors for the systems are

$\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}, \boldsymbol{y}$ global coordinate system,

$\hat{\boldsymbol{i}}, \hat{\boldsymbol{j}}, \hat{\boldsymbol{k}}$ ray coordinate system,

$\hat{\boldsymbol{t}}, \hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}$ vertex coordinate system,

$\hat{\boldsymbol{s}}, \hat{\boldsymbol{p}}, \hat{\boldsymbol{n}}$ surface coordinate system.

The ray, vertex, and surface matrices are described in terms of unit vectors with respect to the global coordinate system. A matrix corresponding to the coordinates systems may be defined in terms of the three column vectors of each system.

$$\mathbf{R}_{gr} = \left[\hat{\boldsymbol{i}}\ \hat{\boldsymbol{j}}\ \hat{\boldsymbol{k}}\right] \qquad \text{ray coordinate matrix,} \tag{11.1}$$

$$\mathbf{R}_{gv} = \left[\hat{\boldsymbol{t}}\ \hat{\boldsymbol{u}}\ \hat{\boldsymbol{v}}\right] \qquad \text{vertex coordinate matrix,}$$

$$\mathbf{R}_{gs} = \left[\hat{\boldsymbol{s}}\ \hat{\boldsymbol{p}}\ \hat{\boldsymbol{n}}\right] \qquad \text{surface coordinate matrix.}$$

The coordinate matrices may be used as rotation matrices to redefine vectors in different coordinate systems. Consider the vector, $\boldsymbol{a}_g$, where we use the subscript, *g*, to indicate the vector is defined in global coordinates. We will use *v*, *r*, and *s* as subscripts to indicate vectors in vertex, ray, and surface coordinates. In vertex coordinates, the same vector would have the identification, $\boldsymbol{a}_v$. The transpose of the coordinate matrix may be used to rotate

$$\boldsymbol{a}_r = \mathbf{R}_{rg}\boldsymbol{a}_g, \tag{11.2}$$

where

$$\mathbf{R}_{rg} = \mathbf{R}_{gr}^{t}. \tag{11.3}$$

The superscripted *t* indicates matrix transpose. The inverse of a rotation matrix is the same as its transpose. The global coordinate system is initialized with the `global` command. The position of elements may be defined on this global coordinates system by specifying the vertex position with the command, `vertex/locate`.

The rotation of the element about the vertex position may be specified with the command, `vertex/rotate`. The vertex location and rotation may be defined absolutely or with reference to the current position of an optical beam. The rotation of the vertex is defined by the user by Euler angles. Specification takes the form

```
vertex/rotate/add kbeam α β γ
vertex/rotate/set α β γ
```

`/add` adds the angles α, β, and γ to the current values and `/set` resets the values. The rotation angles α, β, and γ apply to x-, y-, and z-axis rotations. These rotations obey the right hand rule. A positive x-rotation is done by rotating in the direction of the fingers of the right hand when the thumb is aligned with the positive x-direction. The rotations are applied in the order x, y, and z. Alternate orders may be used by calling `vertex/rotate` several times. For example,

```
vertex/rotate/absolute kbeam 0 0 γ
vertex/rotate/absolute kbeam 0 β 0
vertex/rotate/absolute kbeam α 0 0
```

These commands cause the rotation order to be z, y, and x. See the GLAD Command Description for more detail on `vertex`.

The rotations are defined to be

$$\mathbf{R}_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}, \tag{11.4}$$

$$\mathbf{R}_y(\beta) = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\alpha \end{bmatrix},$$

$$\mathbf{R}_z(\gamma) = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The composite rotation of all three angles (in the order x, y, and z) is

$$\mathbf{R}_{xyz} = \begin{bmatrix} \cos\gamma\cos\beta & \cos\gamma\cos\beta\sin\alpha - \sin\gamma\cos\alpha & \cos\gamma\sin\beta\cos\alpha + \sin\gamma\sin\alpha \\ \sin\gamma\cos\beta & \sin\gamma\sin\beta\sin\alpha + \cos\gamma\cos\alpha & \sin\gamma\sin\beta\cos\alpha - \cos\gamma\sin\alpha \\ -\sin\beta & \cos\beta\sin\alpha & \cos\beta\cos\alpha \end{bmatrix}. \tag{11.5}$$

The `mirror/global` command will correctly calculate the change in the beam for arbitrarily large angles. The commands `clap` and `obs` may be applied at nonnormal angles.

11.1.1 Code V Translation

The global commands make use of beam selection under the command, `beams`. With `beams`, we may selectively turn on or off all or individual beams for all subsequent positioning and mirror/global commands. This facilitates defining a specific optical path of a complex task by allowing all components in the path to be defined in sequence. GLAD specifies the rotation and translation of the vertex. Traditional lens design programs specify coordinate breaks. The rotation directions may also be different. For example in Code V, a product of Optical Research Associates, the rotation Euler angles are defined by ADE, BDE, and CDE. These are coordinate break rotations. The rotation of a component located after such coordinate rotations is the inverse of the net coordinate rotation. Let us define the Code V matrices in primed notation so that the transformation from Code V coordinate break rotation operations may be written in terms of GLAD, right-hand role object rotations:

$$\text{Code V rotations } \mathbf{R}'_{CDE}(\gamma)\mathbf{R}'_{BDE}(\beta)\mathbf{R}'_{ADE}(\alpha) \;\Leftrightarrow\; \text{GLAD system } \mathbf{R}_x(\alpha)\mathbf{R}_y(\beta)\mathbf{R}_z(-\gamma). \tag{11.6}$$

This object rotation may be written as an inverse of the ordinary order of x-, y-, and z-rotations

$$\mathbf{R}_x(\alpha)\mathbf{R}_y(\beta)\mathbf{R}_z(-\gamma) = [\mathbf{R}_z(\gamma)\mathbf{R}_y(\beta)\mathbf{R}_x(-\alpha)]^{-1}. \tag{11.7}$$

We may use the GLAD command,

`vertex/rotate/set/reverse` -α -β γ

which applies the inverse of the ordinary forward GLAD rotation. The inverse GLAD rotation is

$$[\mathbf{R}_z\mathbf{R}_y\mathbf{R}_x]^{-1} = \mathbf{R}_x^{-1}\mathbf{R}_y^{-1}\mathbf{R}_z^{-1}. \tag{11.8}$$

11.2 Chief Ray Propagation

The propagation of a beam through an optical configuration may be described by following the chief ray. The chief ray in GLAD is defined to be the path of the center of the beam array irrespective of the actual complex amplitude distribution in the array. The movement of the chief ray is described by geometrical optics.

We begin by defining several vectors. Since each is in global coordinates, we will omit the g-suffix.

$\boldsymbol{r}$ chief ray position,

$\hat{\boldsymbol{k}}$ chief ray direction,

$\boldsymbol{v}$ vertex location,

$\boldsymbol{s}$ surface intercept of chief ray.

Movement along a ray is defined as a change in the physical path length (PPL). If the PPL is q, then the vector equation is

$$\hat{\boldsymbol{r}}_1 = \hat{\boldsymbol{r}}_0 + q\hat{\boldsymbol{k}}, \tag{11.9}$$

where $\hat{\boldsymbol{r}}_0$ is the starting position and $\hat{\boldsymbol{r}}_1$ is the final position. GLAD does not use the convention of assigning negative index after reflection from a mirror. The index of refraction is always positive. Consequently, only positive values of q are physically meaningful.

GLAD encourages the user to make only forward movements. Redirection of the beam should be made with the `mirror` command. To facilitate the use of index of refraction, we use both optical and physical path length information. The total optical path length (OPL) along the ray through various media is

$$\text{OPL} = \sum_{i=1}^{N} n_i q_i, \tag{11.10}$$

where n_i are the refractive indices and q_i are the distances traveled along the ray in the ith media.

11.2.1 Chief Ray Surface Intercept

We first find the point of closest approach to the vertex of the conic. This approach is more numerically stable than finding the distance to the vertex tangent plane, since the tangent plane may rotate so that it is far from normal to the ray. Let $\boldsymbol{\Delta}_{r_0 v}$ be the translation vector from a point on the ray, $\boldsymbol{r}_0$ to the vertex of the surface,

$$\boldsymbol{\Delta}_{r_o v} = \boldsymbol{v} - \boldsymbol{r}_0. \tag{11.11}$$

The propagation distance, q_1, to the point of closest approach is

$$q_1 = \Delta_{r_o v} \cdot \hat{\boldsymbol{k}}. \tag{11.12}$$

The new vector position is

$$\boldsymbol{r}_1 = \boldsymbol{r}_0 + q_1 \hat{\boldsymbol{k}}. \tag{11.13}$$

The equation for propagation along the ray to the interception point on the surface is

$$\boldsymbol{r}_2 = \boldsymbol{r}_1 + q_2 \hat{\boldsymbol{k}}, \tag{11.14}$$

where $\boldsymbol{r}_2$ is the chief ray intercept on the surface. The incremental distance from the point of closest approach to the vertex to the surface is q_2. See Fig. 11.2.

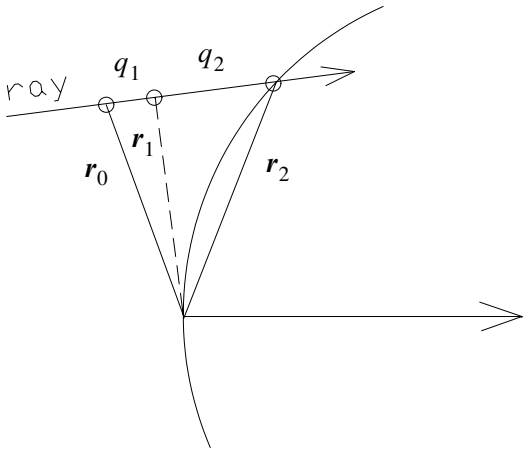


Fig. 11.2. Parameters for ray propagation.

The solution for q_2 is described below. It is convenient to work in vertex coordinates for this problem

$$\Delta_{r_1 v} = \boldsymbol{v} - \boldsymbol{r}_1. \tag{11.15}$$

The coordinates of the optical element are expressed in vertex coordinates. We will use $\hat{\boldsymbol{t}}$, $\hat{\boldsymbol{u}}$, and $\hat{\boldsymbol{v}}$ as vertex unit vectors,

$$\boldsymbol{t} = \Delta_{r_1 v} \cdot \hat{\boldsymbol{t}},\ \boldsymbol{u} = \Delta_{r_1 v} \cdot \hat{\boldsymbol{u}}, \text{ and } \boldsymbol{v} = \Delta_{r_1 v} \cdot \hat{\boldsymbol{v}}. \tag{11.16}$$

The equation for a sphere, in vertex coordinates, for position vector $\boldsymbol{r}_2$ is

$$r_2^2 + u_2^2 + (v_2 - R)^2 = R^2, \tag{11.17}$$

where R is the radius of the surface (also defined in vertex coordinates), the vertex of the spheres is set to (0,0,0), and (t_2, u_2, v_2) is the point of intersection

$$\boldsymbol{r}_2 \cdot \boldsymbol{r}_2 - 2Rz_2 = 0. \tag{11.18}$$

Similarly, the equation for a conic, in vertex coordinates, for position vector $\boldsymbol{r}_2$

$$r_2^2 + u_2^2 + (v_2 - R)^2 + \kappa v_2^2 = R^2, \tag{11.19}$$

$$\boldsymbol{r}_2 \cdot \boldsymbol{r}_2 - 2Rz_2 - \kappa v_2^2 = 0. \tag{11.20}$$

Since the equation for ray propagation is,

$$\boldsymbol{r}_2 = \boldsymbol{r}_1 + q_s \hat{\boldsymbol{k}}. \tag{11.21}$$

This gives us

$$r_2^2 = r_1^2 + q_s^2 + 2q_s(\boldsymbol{r}_1 \cdot \hat{\boldsymbol{k}}) = r_1^2 + q_s^2. \tag{11.22}$$

where $\boldsymbol{r}_1 \cdot \hat{\boldsymbol{k}} = 0$, since $\boldsymbol{r}_1$ is at the point of closest approach and therefore perpendicular to the ray

$$q_s^2 + \boldsymbol{r}_1 \cdot \boldsymbol{r}_1 - 2Rv_2 + \kappa v_2^2 = 0. \tag{11.23}$$

We also have

$$\kappa v_2^2 = \kappa(v_1^2 + 2q_s v_1 k_v + q_s^2 k_v^2), \tag{11.24}$$

$$q_s^2 + \boldsymbol{r}_1 \cdot \boldsymbol{r}_1 - 2Rv_2 + \kappa v_2^2 - 2Rk_v q_s + \kappa(v_1^2 + 2q_s v_1 k_v + q_s^2 k_v^2) = 0, \tag{11.25}$$

$$(1 + \kappa k_v^2) q_s^2 + 2k_v(kv_1 - R) q_s + \boldsymbol{r}_1 \cdot \boldsymbol{r}_1 - 2Rv_1 + \kappa v_1^2 = 0. \tag{11.26}$$

The above equation has the form,

$$Aq_s^2 - 2Bq_s + D = 0, \tag{11.27}$$

where

$$A = 1 + \kappa k_v^2 , \tag{11.28}$$

$$B = k_z(R - \kappa v_1) , \tag{11.29}$$

$$D = -Rv_1 + (t_1 + u_1^2 + (1 + \kappa)v_1^2) , \tag{11.30}$$

$$q_s = \frac{B - I\sqrt{B^2 - AD}}{A} = \frac{D}{B + I\sqrt{B^2 - AD}} , \tag{11.31}$$

where $I = \pm 1$.

Table. 11.1. Selection of quadratic form giving best numeric accuracy.

	$B < 0$	$B > 0$
negative root	$q_s = \dfrac{D}{B - \sqrt{B^2 - AD}}$	$q_s = \dfrac{B + I\sqrt{B^2 - AD}}{A}$
positive root	$q_s = \dfrac{B - \sqrt{B^2 - AD}}{A}$	$q_s = \dfrac{D}{B + \sqrt{B^2 - AD}}$

In general, conic surfaces have two intersection points. In GLAD, we choose the root with the smallest positive value for $q_1 + q_2$, unless `asi` has been chosen in the `mirror/global` command. This results in the smallest forward propagation step. Backward propagation to a surface is not allowed, unlike ray trace codes. If neither root results in a positive $q_1 + q_2$ sum, a ray error is issued. This implies that all of the conic surface is behind the ray. The user should consider the position of the beams at the point in the GLAD command sequence at which the vertex is defined. If `asi` is selected, the root giving the largest $q_1 + q_2$ sum is chosen. If this distance is negative, a ray failure error occurs.

Having calculated the optical path length to the surface intercept, a call to `prop` is issued for the $q_1 + q_2$. `prop` is similar to the `dist` command except that propagation is always forward along the chief ray and the global coordinates are altered by `prop`.

By propagating to the chief ray intercept, all of the diffraction effects are accounted for in the region of the transverse distribution near the chief ray. Regions of the transverse distribution distant form the chief ray have the correct propagation distance, but the aberrations due to reflection from the optical component will, in general, occur earlier or later in the propagation step than they should.

11.2.2 Calculation of Local Surface Properties

To calculate the surface coordinate system it is first necessary to determine the surface normal. In the special case of a conic, the normal to the surface, $\hat{\boldsymbol{n}}$, is calculated by taking the negative of the surface gradient

$$\nabla f = \begin{bmatrix} \frac{t}{R} \\ \frac{u}{R} \\ 1-(1+\kappa)\frac{v}{R} \end{bmatrix}, \tag{11.32}$$

$$\hat{\boldsymbol{n}} = -\frac{\nabla f}{|\nabla f|}. \tag{11.33}$$

GLAD uses the convention that the surface normal points toward the center of curvature of the surface.

In the more general case, we may have a surface defined by the parametric equation for position vector r such that

$$\boldsymbol{r}(x, y, z) = x\hat{\boldsymbol{i}} + y\hat{\boldsymbol{j}} + z\hat{\boldsymbol{k}}. \tag{11.34}$$

The tangent of the surface in the x- and y-directions are

$$\frac{d\boldsymbol{r}}{dx} = \hat{\boldsymbol{i}} + \frac{\partial}{\partial x} f(x, y)\hat{\boldsymbol{k}}, \; \hat{\boldsymbol{t}}_x = \frac{\hat{\boldsymbol{i}} + \frac{\partial}{\partial x} f(x, y)\hat{\boldsymbol{k}}}{\left|\hat{\boldsymbol{i}} + \frac{\partial}{\partial x} f(x, y)\hat{\boldsymbol{k}}\right|}, \tag{11.35}$$

$$\frac{d\boldsymbol{r}}{dy} = \hat{\boldsymbol{j}} + \frac{\partial}{\partial y} f(x, y)\hat{\boldsymbol{k}}, \; \hat{\boldsymbol{t}}_y = \frac{\hat{\boldsymbol{j}} + \frac{\partial}{\partial y} f(x, y)\hat{\boldsymbol{k}}}{\left|\hat{\boldsymbol{j}} + \frac{\partial}{\partial y} f(x, y)\hat{\boldsymbol{k}}\right|}. \tag{11.36}$$

The surface normal is found from

$$\hat{\boldsymbol{n}} = \frac{\hat{\boldsymbol{t}}_x \times \hat{\boldsymbol{t}}_y}{\left|\hat{\boldsymbol{t}}_x \times \hat{\boldsymbol{t}}_y\right|}. \tag{11.37}$$

11.2.3 Calculation of the Surface Coordinate System

Having found $\hat{\boldsymbol{n}}$ and knowing the chief ray direction $\hat{\boldsymbol{k}}$, we may define the surface coordinate system. Two tangent rays on the surface may be defined, $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{p}}$. These correspond to the s- and p-directions used for polarization calculations. The vector $\hat{\boldsymbol{p}}$ is in the plane of reflection and $\hat{\boldsymbol{s}}$ is perpendicular to the plane of reflection, as shown in Figure 11.3,

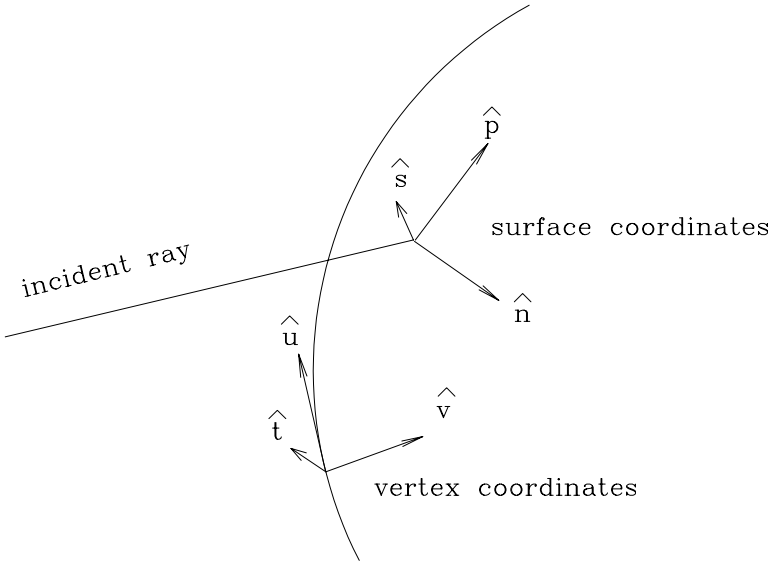


Fig. 11.3. Illustration of surface coordinates.

$$\hat{\boldsymbol{s}} = \frac{\hat{\boldsymbol{k}} \times \hat{\boldsymbol{n}}}{\left| \hat{\boldsymbol{k}} \times \hat{\boldsymbol{n}} \right|}, \tag{11.38}$$

$$\hat{\boldsymbol{p}} = \hat{\boldsymbol{n}} \times \hat{\boldsymbol{s}} \ . \tag{11.39}$$

The surface coordinate system is

$$\mathbf{R}_{gs} = \left[\hat{\boldsymbol{s}} \ \hat{\boldsymbol{p}} \ \hat{\boldsymbol{n}} \right], \tag{11.40}$$

where the parenthesis indicate concatenation of column vectors into a matrix. When the beam strikes the surface at normal incidence, the cross product of $\hat{\boldsymbol{n}}$ and $\hat{\boldsymbol{k}}$ is zero and $\hat{\boldsymbol{s}}$ is not defined. However, at normal incidence, it is not necessary to distinguish between the s- and p-vectors. In practice, when the cross product of Eq. (11.38) drops below a threshold value, GLAD assigns $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{p}}$ to be identical to $\hat{\boldsymbol{i}}$ and $\hat{\boldsymbol{j}}$.

11.3 Reflection

The ray matrix after reflection may be calculated by reversing the sign of the components of the matrix which are parallel to the surface normal. The ray matrix in global coordinates is $\mathbf{R}_{gr}$. In surface coordinates the ray matrix is

$$\mathbf{R}_{sr} = \mathbf{R}_{sg}\mathbf{R}_{gr}. \tag{11.41}$$

We may write the components of $\mathbf{R}_{sr}$ as

$$\mathbf{R}_{sr} = \begin{bmatrix} i_s & i_p & i_n \\ i_s & j_p & j_n \\ k_s & k_p & k_n \end{bmatrix}. \tag{11.42}$$

The reflection operation changes the sign of the n-components

$$\mathbf{R}_{sr} = \begin{bmatrix} i_s & i_p & i_n \\ i_s & j_p & j_n \\ -k_s & -k_p & -k_n \end{bmatrix}. \tag{11.43}$$

The reflected ray matrix in global coordinates is

$$\mathbf{R}_{gr} = \mathbf{R}_{gs}\mathbf{R}_{sr}. \tag{11.44}$$

An alternative form of the reflection equation, which is independent of the coordinate system, is

$$\boldsymbol{i}' = \hat{\boldsymbol{i}} - 2(\hat{\boldsymbol{i}} \cdot \hat{\boldsymbol{n}})\hat{\boldsymbol{n}}, \tag{11.45a}$$

$$\boldsymbol{j}' = \hat{\boldsymbol{j}} - 2(\hat{\boldsymbol{j}} \cdot \hat{\boldsymbol{n}})\hat{\boldsymbol{n}}, \tag{11.45b}$$

$$\boldsymbol{k}' = \hat{\boldsymbol{k}} - 2(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}})\hat{\boldsymbol{n}}. \tag{11.45c}$$

Equation (11.45a) shows that reflection consists of reversing the component in the $\hat{\boldsymbol{n}}$-direction.

11.4 Beam Print on Surface

The beam print on the optical surface may be determined by first finding the surface coordinate system. The beam print on the optical surface is, of course, the same for refractive or reflective surfaces because the beam print is determined only by the incident beam and the surface normal, not by the exiting ray. The incident beam is determined by the ray matrix,

$$\mathbf{K} = [\hat{\boldsymbol{i}}, \hat{\boldsymbol{j}}, \hat{\boldsymbol{k}}]. \tag{11.46}$$

The surface coordinate matrix is determined by the surface normal and the incident ray $\hat{\boldsymbol{k}}$ and Eq. (11.38). If a circular beam is projected onto the surface the magnification in the s- and p-directions is (see Fig. 11.4)

$$M_s = 1, M_p = \frac{1}{|\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}}|}. \tag{11.47}$$

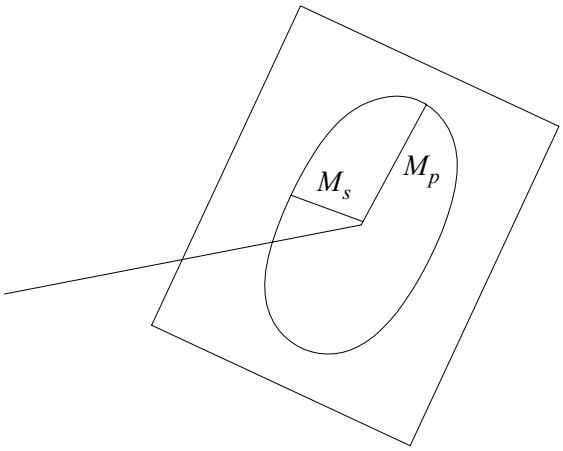


Fig. 11.4. Magnification in the s- and p-directions.

Given an arbitrary azimuthal orientation of the incident ray matrix **K**, the projections of the $\hat{\boldsymbol{i}}$ and $\hat{\boldsymbol{j}}$ vectors onto the surface are

$$M_i = \frac{1}{\left|\hat{\boldsymbol{i}} \times \hat{\boldsymbol{n}}\right|}, \; M_j = \frac{1}{\left|\hat{\boldsymbol{j}} \times \hat{\boldsymbol{n}}\right|}. \tag{11.48}$$

For a more detailed treatment of magnification of the beam print on tilted surfaces see Section 11.6.4.

11.5 Refraction

To compute the ray matrix after refraction we use the refracted direction vector $\hat{\boldsymbol{k}}'$,

$$\hat{\boldsymbol{k}}' = \frac{1}{n'}(n\hat{\boldsymbol{k}} + \Gamma\hat{\boldsymbol{n}}), \tag{11.49}$$

$$\Gamma = n'\cos(I') - n\cos(I), \tag{11.50}$$

$$n'\cos(I') = +\sqrt{n'^2 - n^2[1 - \cos^2(I)]}. \tag{11.51}$$

The incident angle is found from the incident ray vectors

$$I = \text{sign}[\text{acos}(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}}), (\hat{\boldsymbol{k}} \times \hat{\boldsymbol{n}}) \cdot \hat{\boldsymbol{i}}]. \tag{11.52}$$

We can use $\hat{\boldsymbol{k}}'$ and $\hat{\boldsymbol{k}}$ to find a ray matrix referenced to the plane of refraction. $\hat{\boldsymbol{s}}$ is perpendicular to the plane of refraction, and $\hat{\boldsymbol{j}}_r$ orthogonal to $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{k}}$ and $\hat{\boldsymbol{j}}'_r$ orthogonal to $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{k}}'$

$$\hat{\boldsymbol{s}} = \frac{\hat{\boldsymbol{k}} \times \hat{\boldsymbol{n}}}{\left|\hat{\boldsymbol{k}} \times \hat{\boldsymbol{n}}\right|} = \frac{\hat{\boldsymbol{k}}' \times \hat{\boldsymbol{n}}}{\left|\hat{\boldsymbol{k}}' \times \hat{\boldsymbol{n}}\right|}, \; \hat{\boldsymbol{j}}_r = \hat{\boldsymbol{k}} \times \hat{\boldsymbol{s}}, \; \hat{\boldsymbol{j}}'_r = \hat{\boldsymbol{k}}'_r \times \hat{\boldsymbol{s}}. \tag{11.53}$$

An incident ray vector $\mathbf{K}_r$ may be defined

$$\mathbf{K}_r = \left[\hat{\boldsymbol{s}}\ \hat{\boldsymbol{j}}_r\ \hat{\boldsymbol{k}}\right],\ \mathbf{K}_r^{'} = \left[\hat{\boldsymbol{s}}\ \hat{j}'_r\ \hat{\boldsymbol{k}}'\right]. \tag{11.54}$$

The new ray matrix, carrying forward any azimuthal rotations contained in the original matrix

$$\mathbf{K}_r^{'} = \mathbf{R}_r\mathbf{K}_r, \tag{11.55}$$

where the refraction rotation matrix is

$$\mathbf{R}_r = \mathbf{K}_r^{'}\mathbf{K}_r^{-1}, \tag{11.56}$$

$$C22_r = R20_s\left[\frac{1}{\cos(I)} - \cos(I)\right] + \frac{1}{4}C22_s\left[\frac{1}{\cos(I)} + 1\right]. \tag{11.57}$$

11.5.1 Exact Ray Tracing

Exact ray tracing is done in the same fashion as the chief ray up to the surface intercept. The first reference surface is the plane perpendicular to the chief ray at the chief ray intercept. The second reference surface is the plane perpendicular to the exiting chief ray. See Fig. 11.5. For each point in the transverse distribution a new starting ray position and direction is determined. We define the vectors

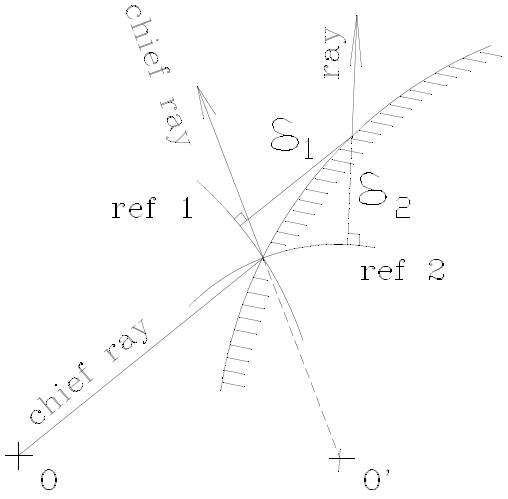


Fig. 11.5. The chief ray intercept determines the vertex of the incident local reference surface. The reflected ray and the local wavefront curvature determine the exiting local reference surface. δ_1 and δ_2 are the optical paths to the surface from the incident and exiting reference surfaces respectively.

$\boldsymbol{r}$ position vector of chief ray,

$\hat{\boldsymbol{k}}$ direction vector of chief ray,

$\boldsymbol{r}_{xy}$ position vector for each x, y point,

$\hat{\boldsymbol{k}}_{xy}$ direction vector for each x, y point.

The position vector is calculated to be

$$\boldsymbol{r}_{xy} = \boldsymbol{r} + x\hat{\boldsymbol{i}} + y\hat{\boldsymbol{j}}, \tag{11.58}$$

$$\hat{\boldsymbol{k}}_{xy} = \frac{\hat{\boldsymbol{k}} - \frac{x}{R_x}\hat{\boldsymbol{i}} - \frac{y}{R_y}\hat{\boldsymbol{j}}}{\left|\hat{\boldsymbol{k}} - \frac{x}{R_x}\hat{\boldsymbol{i}} - \frac{y}{R_y}\hat{\boldsymbol{j}}\right|}, \tag{11.59}$$

where xi are the local coordinates of the transverse beam distribution and R_x and R_y are the phase bias radii. In this approach, the ray slopes are determined by the phase bias radii. Exact slopes could be taken, but the errors with the simpler assumptions are very small. We shall need the following vectors

$\boldsymbol{s}$ chief ray intercept,

Δ_{xy} distance from chief ray surface intercept to current ray intercept in global coordinates.

The reflected direction vector for the current ray, indicated by primes on the coordinates, is

$$\boldsymbol{k'}_{xy} = \hat{\boldsymbol{k}}_{xy} - 2(\hat{\boldsymbol{k}}_{xy} \cdot \hat{\boldsymbol{n}})\hat{\boldsymbol{n}}. \tag{11.60}$$

The vector between the chief ray intercept and the intercept of the current ray being traced is

$$\Delta_{xy} = \boldsymbol{r}_{xy} - \boldsymbol{s}. \tag{11.61}$$

Consider reference planes perpendicular to the chief ray and reflected chief ray. Let δ_1 be the distance from the chief ray reference plane to the surface measured along the chief ray and δ_2 be the distance from the surface intercept to the reflected chief ray reference plane measured along the reflected chief ray. Then

$$\delta_1 = \frac{\Delta_{xy} \cdot \hat{\boldsymbol{k}}}{\hat{\boldsymbol{k}}_{xy} \cdot \hat{\boldsymbol{k}}}, \; \delta_2 = \frac{\Delta_{xy} \cdot \boldsymbol{k'}}{\hat{\boldsymbol{k}}'_{xy} \cdot \boldsymbol{k'}}, \tag{11.62}$$

where primed quantities indicated imaged quantities. The transverse intercepts on the reflected chief ray reference plane are found from δ_2

$$x' = (\boldsymbol{\Delta}_{xy} + \delta_2 \boldsymbol{k}') \cdot \hat{\boldsymbol{i}}', \; y' = (\boldsymbol{\Delta}_{xy} + \delta_2 \boldsymbol{k}') \cdot \hat{\boldsymbol{j}}', \tag{11.63}$$

where the unit vectors of the reflected system are found with Eq. (11.60). The complex amplitude is transferred from the first reference plane to the second by

$$a(x', y') = a(x, y) e^{jk(\delta_1 + \delta_2)}. \tag{11.64}$$

In general, the point (x', y') will not lie on a grid point in the new array. The value must be interpolated to the nearest neighboring grid points.

Figure 11.5 shows a ray emanating from a point on Reference Surface 1. Such a ray will exhibit a slightly different wavefront slope as a function of the aberration in the beam. The optical path length from the Reference Surface 1 to Reference Surface 2 is relatively insensitive to the exact path of the ray. This is a consequence of Fermat's Principle which states that the optical path length along an optical ray is stationary, i.e., closely neighboring paths have essentially the same length.

11.6 Mirror Systems, Image Rotations, Image Parity, Common Prisms

Plane mirrors are used to pipe the optical beam through a system, to steer the beam to a specific direction, and to provide image rotation. It is commonly said that a mirror causes a left-to-right reversal. This is not strictly correct: a single mirror cause a front-to-back reversal which has a reversal of parity associated with it. There are mirror systems which actually do cause a left-to-right reversal, such as the K-mirror, which will be discussed. In elementary optical texts it is common to illustrate the effect of mirrors by showing how the letter "R" is imaged. This simple concept works well for one mirror and adequately for a simple two-mirror system, but becomes unworkable for more complex systems. The mirror matrix method used in this section is much more powerful and has the advantage that it is readily implemented on a computer. The mirror matrix also has the advantage that all mirror systems having the same mirror matrix will have the same behavior. In many cases the mirror matrix may be determined by inspection.

11.6.1 Mirror Matrices

For a ray direction vector of $\hat{\boldsymbol{k}}$, the equation for reflection from a mirror is (see Fig. 11.6.) we have the equation from Eq. (11.45c)

$$\hat{\boldsymbol{k}}_2 = \hat{\boldsymbol{k}}_1 - 2(\hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{n}})\hat{\boldsymbol{n}}. \tag{11.65}$$

In Dirac notation, $\langle a|$ is a bra and $|a\rangle$ is a ket, $\langle a|b\rangle$ is an inner product

$$\text{inner product } \langle a|b\rangle = a^t b = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = a_1 b_1 + a_2 b_2 + a_3 b_3, \tag{11.66}$$

$|a\rangle\langle b|$ is an outer product

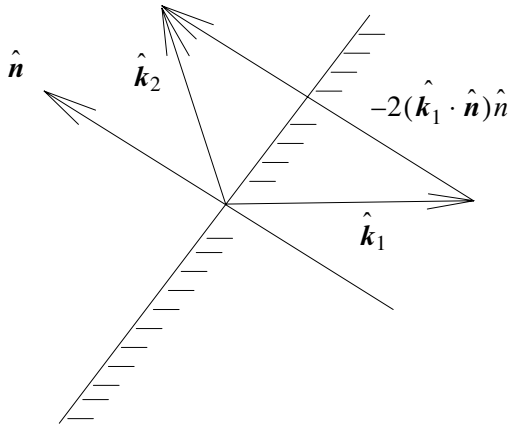


Fig. 11.6. Reflection of k-vector.

$$\text{outer product } |a\rangle\langle b| \;=\; ab^t \;=\; \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix} \;=\; \begin{bmatrix} a_1 b_1 & a_1 b_2 & a_1 b_3 \\ a_2 b_1 & a_2 b_2 & a_2 b_3 \\ a_3 b_1 & a_3 b_2 & a_3 b_3 \end{bmatrix}. \tag{11.67}$$

We can write the equations for incident and reflected k-vectors in terms of the surface normal of the mirror from Eq. (11.65) in bra-ket notation as

$$|k_2\rangle \;=\; |k_1\rangle - 2\langle k_1|n\rangle|n\rangle\,. \tag{11.68}$$

Reversing the order of multiplication of the two elements of the second term on the right side and also reversing the order of terms in the inner product, we have

$$|k_2\rangle \;=\; |k_1\rangle - 2|n\rangle\langle n|k_1\rangle \;=\; [\mathbf{I} - 2|n\rangle\langle n|]|k_1\rangle\,. \tag{11.69}$$

Where **I** is the identity matrix. We have the relationship

$$\mathbf{I} \;=\; \sum_{l\,=\,1}^{3} |l\rangle\langle l|\,. \tag{11.70}$$

From Eq. (11.69) we can define the mirror matrix

$$\mathbf{M} \;=\; \mathbf{I} - 2|n\rangle\langle n|\,. \tag{11.71}$$

Given the surface normal $\hat{\boldsymbol{n}}$, we may compute the mirror matrix using the definition of the outer product from Eq. (11.67).

The Law of Reflection may now be written in terms of the mirror matrix

$$|k_2\rangle = \mathbf{M}|k_1\rangle \text{ or } \hat{\boldsymbol{k}}_2 = \mathbf{M}\hat{\boldsymbol{k}}_1 . \tag{11.72}$$

A series of reflections is calculated by successive mirror matrix multiplications. For example

$$\hat{\boldsymbol{k}}_4 = \mathbf{M}_3\mathbf{M}_2\mathbf{M}_1\hat{\boldsymbol{k}}_1 = \mathbf{M}_t\hat{\boldsymbol{k}}_1 , \tag{11.73}$$

where

$$\mathbf{M}_t = \prod_{i=3}^{1} \mathbf{M}_i . \tag{11.74}$$

The mirror matrix is derived from the mirror normal vector $\hat{\boldsymbol{n}}$. For example a mirror with a surface normal aligned with the z-axis is

$$\hat{\boldsymbol{n}}_z = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} . \tag{11.75}$$

We form the mirror matrix $\mathbf{M}_z$ for this surface normal as follows (see Fig. 11.7):

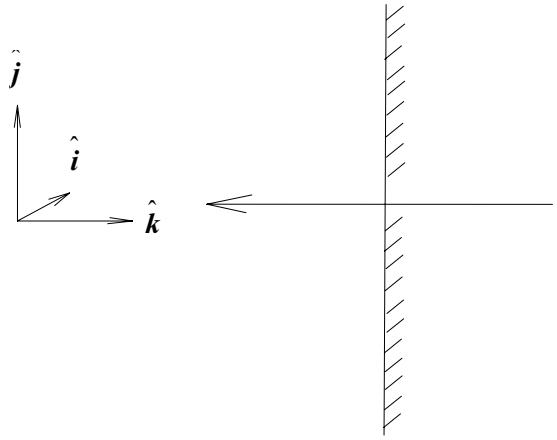


Fig. 11.7. Mirror matrix $\mathbf{M}_z$.

$$|n_z\rangle\langle n_z| = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} , \tag{11.76}$$

$$\mathbf{M}_z = \mathbf{I} - 2|n_z\rangle\langle n_z| = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - 2\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \tag{11.77}$$

The mirror matrices for mirrors with normals in the x- and y-directions are

$$\mathbf{M}_x = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{M}_y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{11.78}$$

The chief ray direction is commonly used in geometrical optics but does not allow the azimuthal beam rotations to be determined. Both the ray direction and azimuthal rotation may be defined by a ray matrix

$$\mathbf{K} = \begin{bmatrix} \hat{\boldsymbol{i}} & \hat{\boldsymbol{j}} & \hat{\boldsymbol{k}} \end{bmatrix} = \begin{bmatrix} i_x & j_x & k_x \\ i_y & j_y & k_y \\ i_z & j_z & k_z \end{bmatrix}. \tag{11.79}$$

A beam matrix aligned with the x-y-z coordinate system has the ray matrix:

$$\mathbf{K} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{11.80}$$

as indicated in Fig. 11.7.

The beam coordinate system may be rotated arbitrarily. The Law of Reflection may be applied to the ray matrix

$$\mathbf{K}_2 = \mathbf{M}\mathbf{K}_1. \tag{11.81}$$

A general condition of reflection of an arbitrarily rotated matrix is illustrated in Fig. 11.8.

A ray matrix may be right or left handed. The matrix defined in Eq. (11.80) is a right-handed coordinate system. The system is right-handed if

$$\hat{\boldsymbol{i}} \times \hat{\boldsymbol{j}} = \hat{\boldsymbol{k}}, \tag{11.82}$$

and left-handed if the cross product is $-\hat{\boldsymbol{k}}$. Left or right handedness may also be called positive or negative parity or chirality. The parity of a matrix may also be determined by calculating the sign of the determinant of the matrix

$$\text{parity} = \det[\mathbf{M}]. \tag{11.83}$$

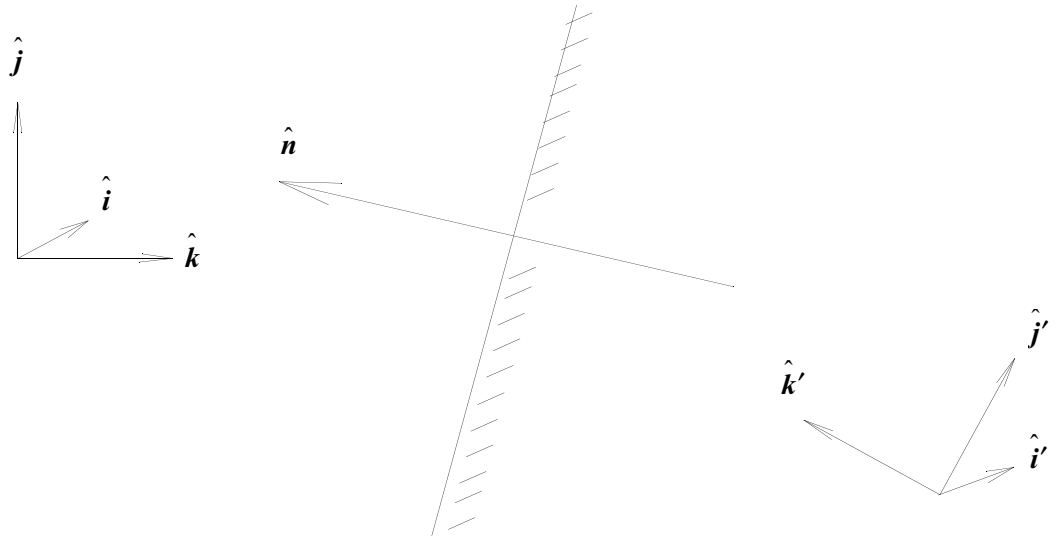


Fig. 11.8. Object and image space characteristics may be defined by the ray matrix.

The matrix of a single mirror flips the parity. For example reflection from a mirror with its surface normal to the z-axis changes the ray matrix to

$$\text{single mirror } \mathbf{M}_z,\ \mathbf{K} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \tag{11.84}$$

The transverse vectors remain the same and the sign of $\hat{\boldsymbol{k}}_z$ is reversed, leading to a left-handed coordinate system. This is sometimes referred to as reversion.

An inverted ray matrix has the form

$$\text{inverted ray matrix } \mathbf{K} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{11.85}$$

as indicated in Fig. 11.9, but has positive parity.

The parity of a matrix product is the product of the parities of the individual matrices. The parity of rotation matrices is always +1 and the parity of any single mirror is -1, so the parity of a system of N mirrors is

$$\text{parity of N mirrors} = (-1)^N. \tag{11.86}$$

If the ray matrix is reflected from a mirror defined by **M** that has been given a rotation described by the rotation matrix **R** the reflected ray matrix has the form

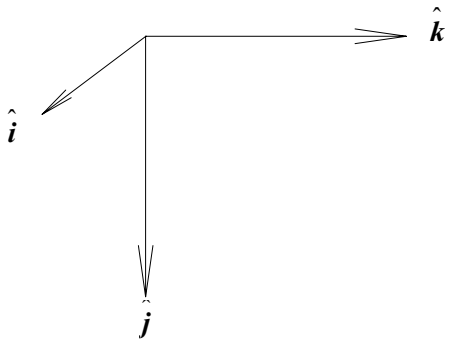


Fig. 11.9. An inverted ray matrix.

$$\mathbf{K}_2 = \mathbf{R}\mathbf{M}\mathbf{R}^{-1}\mathbf{K}_1 \tag{11.87}$$

This is called a similarity transformation and the rotation matrix **R** must be a real unitary matrix which has the properties $\det[\mathbf{R}]$ and $\mathbf{R}^{-1} = \mathbf{R}^t$.

Rotation matrices are described by Eq. (11.4) and (11.5). Mirrors having the orthogonal elementary x-, y-, or z-rotations lend themselves to simple analysis. Some of the properties of these systems are shown in Table 11.2.

Table. 11.2. Effects of rotations on x-, y-, and z-systems, method of cofactors.

x-system	y-system	z-system
$\mathbf{M}_x = \begin{bmatrix} \pm 1 & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix}$	$\mathbf{M}_y = \begin{bmatrix} a_{11} & 0 & a_{13} \\ 0 & \pm 1 & 0 \\ a_{31} & 0 & a_{33} \end{bmatrix}$	$\mathbf{M}_z = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & \pm 1 \end{bmatrix}$
cofactor $\mathbf{M}_{11} = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$	cofactor $\mathbf{M}_{22} = \begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix}$	cofactor $\mathbf{M}_{22} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$
for $\det[\mathbf{M}_{11}] = 1$	for $\det[\mathbf{M}_{22}] = 1$	for $\det[\mathbf{M}_{33}] = 1$
$\mathbf{R}_x(\alpha)\mathbf{M}_x\mathbf{R}_x(-\alpha) = \mathbf{M}_x$	$\mathbf{R}_y(\beta)\mathbf{M}_y\mathbf{R}_y(-\beta) = \mathbf{M}_y$	$\mathbf{R}_z(\gamma)\mathbf{M}_z\mathbf{R}_z(-\gamma) = \mathbf{M}_z$
for $\det[\mathbf{M}_{11}] = -1$	for $\det[\mathbf{M}_{22}] = -1$	for $\det[\mathbf{M}_{33}] = -1$
$\mathbf{R}_x(\alpha)\mathbf{M}_x\mathbf{R}_x(-\alpha) = \mathbf{R}_x(2\alpha)\mathbf{M}_x$	$\mathbf{R}_y(\beta)\mathbf{M}_y\mathbf{R}_y(-\beta) = \mathbf{R}_y(2\beta)\mathbf{M}_y$	$\mathbf{R}_z(\gamma)\mathbf{M}_z\mathbf{R}_z(-\gamma) = \mathbf{R}_z(2\gamma)\mathbf{M}_z$

Consider the arrangement of three mirrors shown in Fig. 11.10 to form a K-mirror system. All three mirrors may be considered to be an x-rotation of a z-mirror. The end mirrors may be given some suitable rotation $\alpha > 90°$. The overall mirror system may be written

$$\mathbf{M}_t = \mathrm{R}_x(-\alpha)\mathrm{M}_z\mathrm{R}_x(\alpha)\mathrm{R}_x\left(\frac{\pi}{2}\right)\mathrm{M}_z\mathrm{R}_x\left(-\frac{\pi}{2}\right)\mathrm{R}_x(\alpha)\mathrm{M}_z\mathrm{R}_x(-\alpha) = \mathrm{R}_x\left(\frac{\pi}{2}\right)\mathrm{M}_z\mathrm{R}_x\left(-\frac{\pi}{2}\right) = \mathrm{M}_y . \qquad (11.88)$$

So the rotation properties of the three mirrors of the K-mirror system are identical to the rotation of the single y-mirror. For an azimuthal rotation, the K-mirror system will rotate the image at twice the rotation rate of the system.

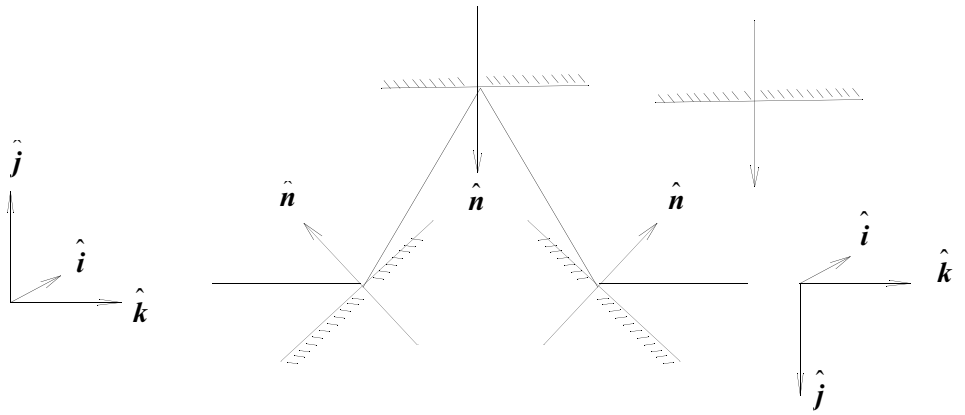


Fig. 11.10. K-mirror system. It is equivalent to the elementary y-mirror system $\mathbf{M}_y$.

We can also construct this matrix directly by inspection. The system passes the beam directly through so the exiting beam is parallel with the input beam so that

$$\hat{k}' = \hat{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \qquad (11.89)$$

The $\hat{i}$ vector will not be changed because all mirrors have x-rotations. The parity is -1 so

$$\hat{j}' = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \qquad (11.90)$$

and the mirror matrix must therefore be,

$$\mathbf{M}_y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad (11.91)$$

The K-mirror system can be used as an image rotator by rotating the whole assembly azimuthally. The image will rotate at twice the angle of the azimuthal rotation. This is obvious as soon as we recognize that the mirror system responds as $\mathbf{M}_y$. See the properties of the y-mirror from Table 11.2. One of the advantages of this is that any system — no matter how complex — will behave according to the behavior of its mirror matrix. The procedures here can be used to analyze any mirror system and prism systems where the refractive surfaces are normal to the beam. This is very common for prism systems since only normal incidence will insure no dispersive effects.

11.6.2 Common Mirrors and Prisms

The following pages illustrate some mirror configurations. The characteristics of the system may often be analyzed by inspection by tracing the k-vector, finding either the i- or j-vector, and deducing the other component from knowledge of the parity.

An important class of systems has the output ray parallel to the input ray. These are called direct vision systems. If the parity is even the image will be insensitive to rotation of the system about the line of sight. If the parity is odd the image will rotate at twice the angle of the azimuthal rotation.

For direct vision prisms, there are often mirror equivalents to the prisms. See MIL-HDBK-141. Prisms which use a surface in both reflection and transmission, such as the Pechan prism, do not have direct mirror equivalents. Although prism systems are heavier than similar mirror systems, they are rigid and, therefore, retain their alignment state under shock and vibration. For example, prisms are always used in binoculars to achieve a rugged instrument.

90º roof prisms have the same reflection properties as a mirror at the same angle put flip the parity. 90º roofs must be made very carefully — any error in the apex angle will result in sever aberration, not just an error in the exiting angle. Roofs other than 90º have two solutions depending on which mirror is struck first. All roofs deviate the incident ray by a fixed angle, independent of rotation about the roof line.

Three-mirror systems are similar to one-mirror systems, except that one has greater control of the configuration. Note carefully the difference between direct vision prisms with odd parity, which are beam rotators and direct vision prism systems with even parity which cause 180º rotation of the image. Figure 11.11 illustrates some elementary mirrors and roof mirrors. Figure 11.12 illustrates some roofs with 45º apex angles and the corner cube. Figure 11.13 illustrates some 90ºdevation systems. Figure 11.14 illustrates some direct vision systems (k-vector exits parallel to input k-vector)..

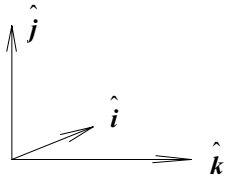
Traditional Terms

invert	180º rotation about the direction of propagation and parity change,
revert	parity change,
erect	image is inverted and reverted, i.e., image is simply rotated 180º,
direct vision	exiting line-of-sight is parallel to entering line-of-sight.

11.6.3 Image Positions

The principles discussed so far may be extended to locate image positions. The position of a mirror may be defined by its normal vector $\hat{\boldsymbol{n}}$ and distance from the coordinate system of q. If we know the position

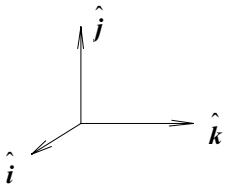
free space $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$



x-mirror $\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

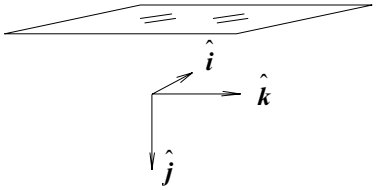
insensitive to x-rotation, 2θ for y- or z-rotations, parity -1





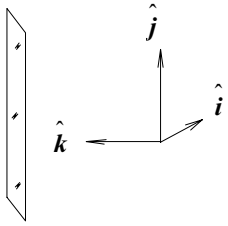
y-mirror $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

insensitive to y-rotation, 2θ for x- or z-rotations, parity -1



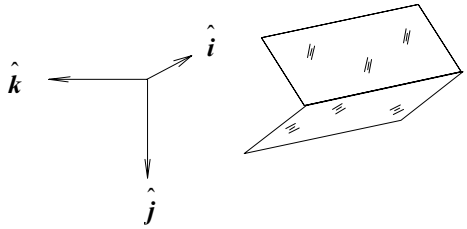
z-mirror $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$

insensitive to z-rotation, 2θ for x- or y-rotations, parity -1



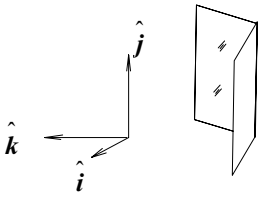
90º x-roof $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$

insensitive to x-rotation, 2θ for y- or z-rotations, parity 1



90º y-roof $\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$

insensitive to y-rotation, 2θ for x- or z-rotations, parity 1



90º z-roof $\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

insensitive to z-rotation, 2θ for x- or y-rotations, parity 1

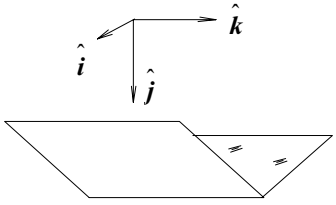
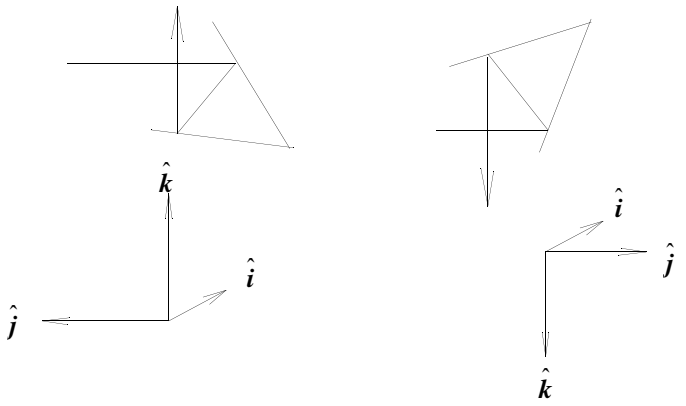


Fig. 11.11. Elementary Mirrors and Roofs.

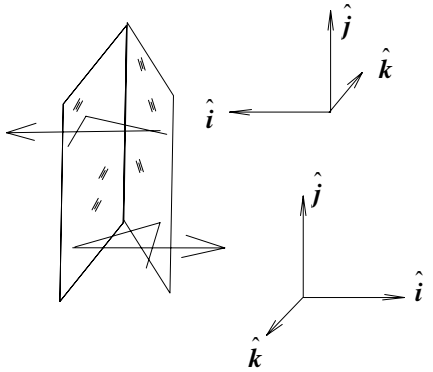
45º x-roof $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & \pm 1 \\ 0 & \mp 1 & 0 \end{bmatrix}$

insensitive to x-rotation, 90º deviation, parity 1



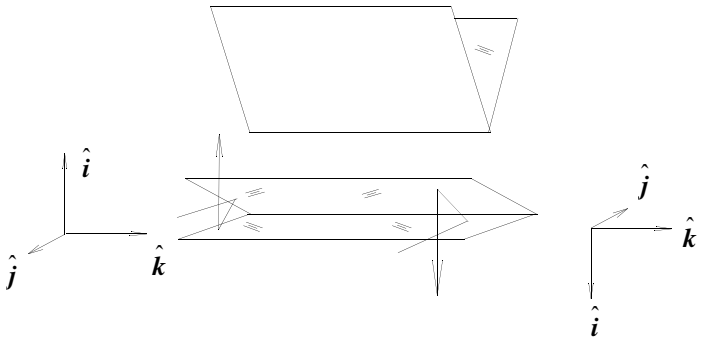
45º y-roof $\begin{bmatrix} 0 & 0 & \pm 1 \\ 0 & 1 & 0 \\ \mp 1 & 0 & 0 \end{bmatrix}$

insensitive to y-rotation, 90º deviation, parity 1



45º z-roof $\begin{bmatrix} 0 & \pm 1 & 0 \\ \mp 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

insensitive to z-rotation, 90º deviation, parity 1



cube corner $\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$

insensitive to all rotations, retroreflects, parity -1

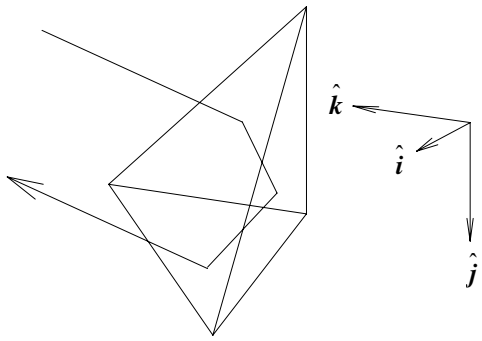
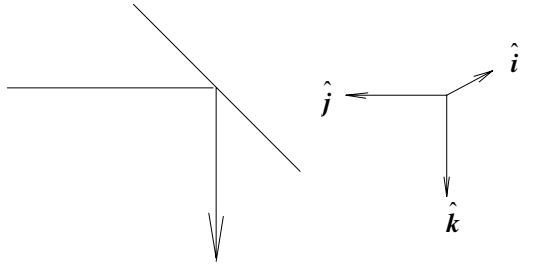


Fig. 11.12. Roofs with 45º apex angles and the corner cube. All roofs other than 90º have two solutions depending on which face the ray hits first. Parity is even, angle of ray deviation is twice the apex angle.

some 90º mirror systems of the form $\begin{bmatrix} \pm1 & 0 & 0 \\ 0 & 0 & \pm1 \\ 0 & \pm1 & 0 \end{bmatrix}$

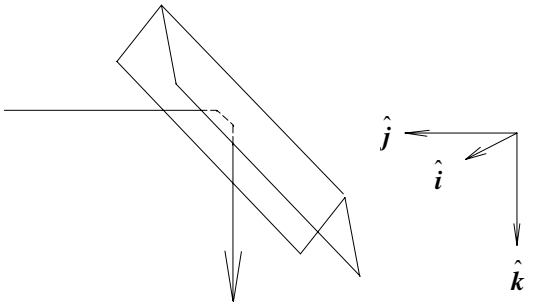
tilted flat mirror $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix}$

flat mirror, 45º prims, Porro, parity 1



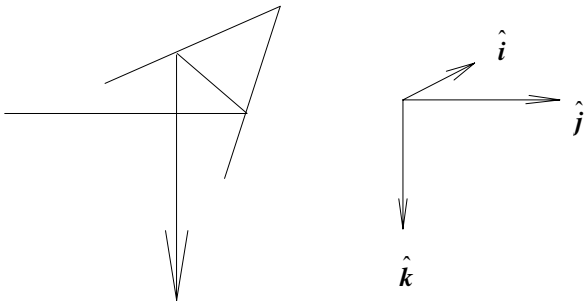
replace flat with roof $\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 0 \end{bmatrix}$

roof replaces flat mirror, Amici prims, parity 1



45º x-roof $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$

pentaprism, 45º x-roof, parity 1



replace one flat mirror with roof $\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$

petaprism with roof on one side, parity -1
depends on whether ray hits flat or roof first

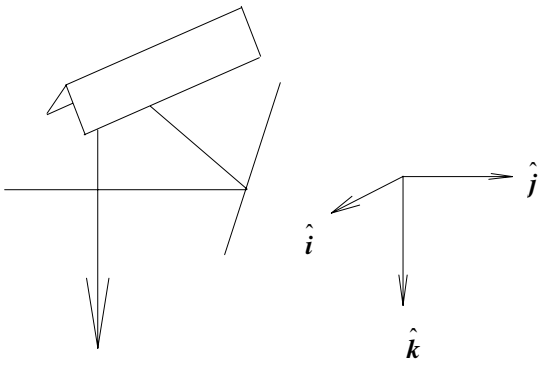
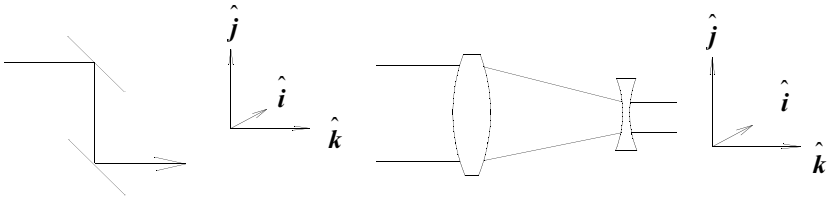


Fig. 11.13. Some 90º deflection systems.

direct division systems, general form $\begin{bmatrix} \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

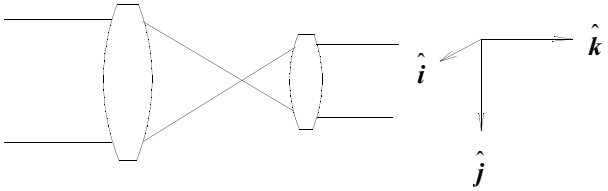
rhomb and galilean telescope, same as free space, parity 1 $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$



Keplarian (or astronomical) telescope, parity 1

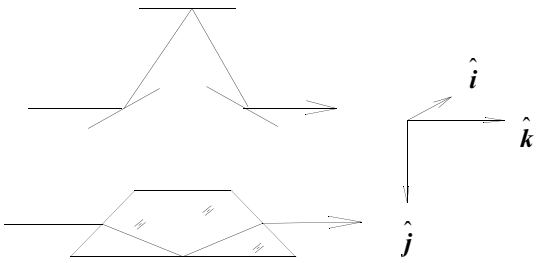
$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

image is inverted and reverted, use erecting prism to erect image



K-mirror, Dove, double Dove, Pechan, reversion prism, useful for rotating an image about the z-axis, parity -1

$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$



erecting prisms, pair of Porro prisms, Abee prism, Leman, parity 1, erects Keplarian telescope image $\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

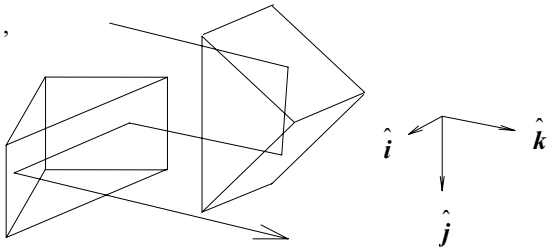


Fig. 11.14. Some direct vision systems.

vector of any point on the surface, e.g., some vector $\boldsymbol{s}$, then the distance to the mirror surface along the perpendicular (the minimum distance) is (see Fig. 11.15)

$$q = \boldsymbol{s} \cdot \hat{\boldsymbol{n}}. \tag{11.92}$$

The vector **s** may be found from the position vector to the point and a position vector r to any point on the mirror $\boldsymbol{r}_M$ (see Fig. 11.16). The i$^{\text{th}}$ optical length is

$$q_i = \boldsymbol{s}_i \cdot \hat{\boldsymbol{n}}_i = (\boldsymbol{r}_{M_1} - \boldsymbol{r}_i) \cdot \hat{\boldsymbol{n}}_i. \tag{11.93}$$

The position of an image point may be designated by the vector $\boldsymbol{r}_0$ and the corresponding image point in the mirror by the vector $\boldsymbol{r}_1$, such that

$$\boldsymbol{r}_1 = \mathbf{M}\boldsymbol{r}_0 + 2q\hat{\boldsymbol{n}}. \tag{11.94}$$

We can find the image point of a series of mirrors by an operator sequence, for example,

$$\boldsymbol{r}_3 = \mathbf{M}_3[\mathbf{M}_2[\mathbf{M}_1\boldsymbol{r}_0 + 2q_1\hat{\boldsymbol{n}}] + 2q_2\hat{\boldsymbol{n}}] + 2q_3\hat{\boldsymbol{n}}. \tag{11.95}$$

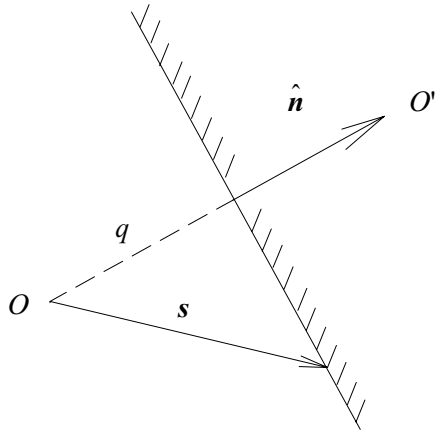


Fig. 11.15. Point O is imaged to O'. ***S.*** is a vector from the object to a point on the mirror. q is the distance to the mirror.

11.6.4 Beam Print on Tilted Optical Surfaces

The beam print on the optical surface may be determined by first finding the surface coordinate system. The beam print on the optical surface is, of course, the same for refractive or reflective surfaces because the

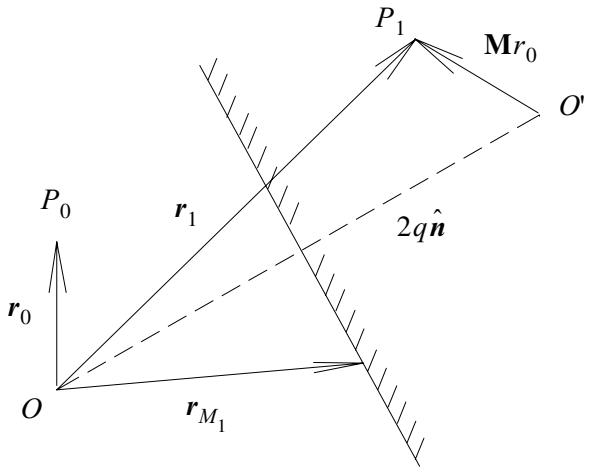


Fig. 11.16. Location vector of P_o is $\boldsymbol{r}_0$. Location vector of P_1 is $\boldsymbol{r}_1 = \mathbf{M}\boldsymbol{r}_0 + 2q\hat{\boldsymbol{n}}$.

beam print is determined only by the incident beam and the surface normal, not by the exiting ray. See Fig. 11.1. The incident beam is determined by the ray matrix Eq. (11.79).

If a circular beam is projected onto the surface, the magnifications in the s- and p-directions are

$$m_s = 1\,,\; m_p = \frac{1}{\left|\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}}\right|}\,. \tag{11.96}$$

Defining the direction of maximum magnification in the surface as a vector, we have

$$\boldsymbol{m}_p^{(0)} = \frac{1}{\left|\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}}\right|}\hat{\boldsymbol{p}}\,. \tag{11.97}$$

Expressing $\hat{\boldsymbol{p}}$ in the beam coordinate system we have

$$\begin{bmatrix} p_i \\ p_j \\ p_k \end{bmatrix} = \mathbf{K}^t\hat{\boldsymbol{p}}\,. \tag{11.98}$$

The azimuthal orientation ϕ is

$$\phi = \tan^{-1}\left(\frac{p_j}{p_i}\right). \tag{11.99}$$

After refraction or diffraction we have the exiting direction vector $\hat{\boldsymbol{k}}_r$. We may now compute an s′-p′-k′ coordinate system with $\boldsymbol{s}'$ perpendicular to the n′-k′ plane:

$$\hat{\boldsymbol{s}}' = \frac{\boldsymbol{k}' \times \hat{\boldsymbol{n}}}{\left|\boldsymbol{k}' \times \hat{\boldsymbol{n}}\right|}, \boldsymbol{p}' = \boldsymbol{n}' \times \boldsymbol{s}'. \tag{11.100}$$

The magnification in the p′-direction is defined by the vector

$$\boldsymbol{m}_p^{(1)} = \left|\hat{\boldsymbol{k}}' \cdot \hat{\boldsymbol{n}}'\right| \hat{\boldsymbol{p}}'. \tag{11.101}$$

In beam coordinates $\hat{\boldsymbol{p}}'$ is

$$\begin{bmatrix} p'_i \\ p'_j \\ p'_k \end{bmatrix} = \mathbf{K}'^t \boldsymbol{p}'. \tag{11.102}$$

The azimuthal orientation ϕ' is found by

$$\phi' = \tan^{-1}\left(\frac{p'_j}{p'_i}\right). \tag{11.103}$$

In the case of refraction and reflection (but not diffraction from a grating in the general case), $\phi = \phi'$ as $\boldsymbol{k}$, $\boldsymbol{n}$, and $\boldsymbol{k}'$ are in a common plane. So the magnification change in the ϕ–direction is

$$m_p = \frac{m_p^{(1)}}{m_p^{(0)}} = \frac{\left|\hat{\boldsymbol{k}}' \cdot \hat{\boldsymbol{n}}\right|}{\left|\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}}\right|}. \tag{11.104}$$

The magnification distortion is applied as

$$\mathbf{M}_{\text{mag}} = \mathbf{R}(-\phi) \begin{bmatrix} m_p & 0 \\ 0 & 1 \end{bmatrix} \mathbf{R}(\phi). \tag{11.105}$$

In the more general case where $\phi \neq \phi'$ that can occur with a plane grating where the grating vector is not in the k-n plane, we must apply the magnifications from beam-to-surface and surface-to-beam in succession

$$\mathbf{M}_{\text{mag}} = \mathbf{R}(-\phi') \begin{bmatrix} m_p^{(1)} & 0 \\ 0 & 1 \end{bmatrix} \mathbf{R}(\phi') \mathbf{R}(-\phi) \begin{bmatrix} m_p^{(0)} & 0 \\ 0 & 1 \end{bmatrix} \mathbf{R}(\phi). \tag{11.106}$$

11.6.5 Accumulating Magnification Over Many Operations

At each surface the magnification takes the form:

$$\mathrm{M}_{\mathrm{mag}} = \mathbf{R}(-\phi)\begin{bmatrix} m & 0 \\ 0 & 1 \end{bmatrix}\mathbf{R}(\phi)\,. \tag{11.107}$$

For a succession of bends at surfaces we have the cumulative effect of all magnifications:

$$\mathrm{M}_{\text{mag-total}} = \prod_{i=1}^{N} \mathbf{R}(-\phi_i)\begin{bmatrix} m_i & 0 \\ 0 & 1 \end{bmatrix}\mathbf{R}(\phi_i)\,. \tag{11.108}$$

After a number of magnifications of arbitrary values and orientations we have a general matrix which may be decomposed into uniform magnification, a distortion term in the x-y direction, and another at some angle ψ

$$\mathrm{M}_{\text{mag-total}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} m_1 & 0 \\ 0 & m_1 \end{bmatrix}\begin{bmatrix} m_2 & 0 \\ 0 & 1/m_2 \end{bmatrix}\mathbf{R}(-\psi)\begin{bmatrix} m_3 & 0 \\ 0 & 1/m_3 \end{bmatrix}\mathbf{R}(\psi)\,. \tag{11.109}$$

The uniform magnification is easily found from the determinant of $\mathrm{M}_{\text{mag-total}}$

$$m_1 = \sqrt{ab - cd}\,, \tag{11.110}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 1/m_1 & 0 \\ 0 & 1/m_1 \end{bmatrix} = \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} = \begin{bmatrix} m_2 & 0 \\ 0 & 1/m_2 \end{bmatrix}\mathbf{R}(-\psi)\begin{bmatrix} m_3 & 0 \\ 0 & 1/m_3 \end{bmatrix}\mathbf{R}(\psi)\,. \tag{11.111}$$

m_2 is solved such that

$$\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}\begin{bmatrix} 1/m_2 & 0 \\ 0 & m_2 \end{bmatrix} \tag{11.112}$$

is symmetric. We solve $m_2 = \sqrt{b'c'}$ such that the result is symmetric

$$\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}\begin{bmatrix} 1/m_2 & 0 \\ 0 & m_2 \end{bmatrix} = \begin{bmatrix} a'' & b'' \\ b'' & c'' \end{bmatrix}, \tag{11.113}$$

$$\begin{bmatrix} a'' & b'' \\ b'' & c'' \end{bmatrix} = \begin{bmatrix} a'\sqrt{\dfrac{c'}{b'}} & \sqrt{b'c'} \\ \sqrt{b'c'} & d'\sqrt{\dfrac{b'}{c'}} \end{bmatrix} = \mathbf{R}(-\psi)\begin{bmatrix} m_3 & 0 \\ 0 & 1/m_3 \end{bmatrix}\mathbf{R}(\psi)\,. \tag{11.114}$$

The final step is to determine m_3 and ψ. They are found by orthogonalizing the symmetric matrix formed by applying m_2 in Eq. (11.113). The amplitude must be adjusted to correct for the change in magnification. The correction factor for amplitude is

$$\text{amplitude correction factor} = \frac{1}{\det[\mathbf{M}_{\text{mag-total}}]}. \tag{11.115}$$

The two distortion terms of Eq. (11.109) are defined with unity value of the determinant. Only the uniform magnification term has non-unity determinant and it exclusively determines the amplitude correction.

11.6.6 Tunnel Diagrams

The optical performance of a prism with normal surface interface angles is the same as that of a plane-parallel plate of equivalent thickness. A prism — no matter how complex — may be represented by a solid block of glass of equivalent optical length. A tunnel diagram may be constructed by "unfolding" the prism. Figures 11.17.a and 11.17b show the appearance of a pentaprism before and after unfolding about the two reflecting surfaces. The reflecting surfaces are TIR so the loss is negligible. A block of the same glass with the same length should have exactly the same optical effects as the pentaprism.

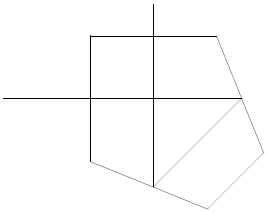


Fig. 11.17.a. Pentaprism produces an exact fold and uses TIR reflection.

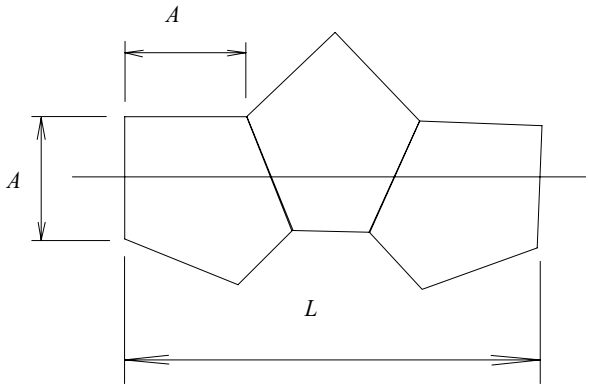


Fig. 11.17b. Pentaprism produces an exact 90º fold and uses TIR reflection.

11.6.7 Reduced Distance

The reduced distance is the length of an air space equivalent of a glass length. The equivalent air space has the same paraxial properties as the glass length represents, i.e., the rise or fall of a ray is the same when traversing the glass path or the equivalent air path. See Fig. 11.18.a. If L is the length in glass, then L/n is the equivalent length in air. The equivalence is strictly a first-order property. A plane parallel piece of glass adds aberrations to an noncollimated beam, while the air path does not. An image, when viewed through glass, looks closer. Similarly adding a glass window to an optical beam moves the image further away. The image is moved by the difference between the true path and the reduced path, i.e., image shift is $(1 - 1/n)L$ — about $L/3$ for low index glass, $n \approx 1.5$. The unfolded pentaprisms has a reduced distance according to the length

of class in the unfolded system. See Fig. 11.18b. The wavelength in glass is λ_0/n where λ_0 is the wavelength in vacuum. The diffraction properties of light depends on the wavelength (not the vacuum wavelength). Diffraction properties will, therefore depend on the reduced length. Aberrations, in contrast, determined by the optical path length differences.

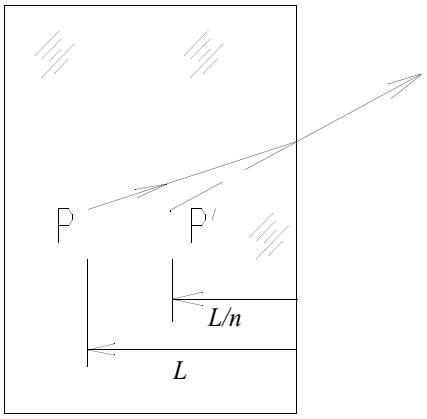


Fig. 11.18.a. Reduced length is the equivalent air path representing propagation through glass. The image shift due to the glass is $(1-1/n)L$.

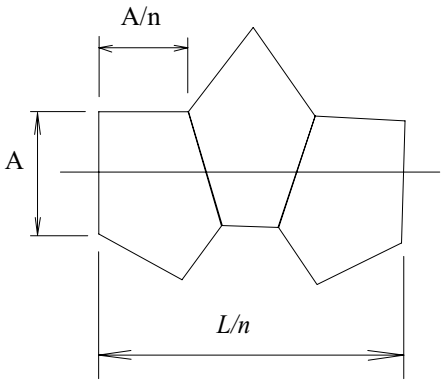


Fig. 11.18b. Reduced tunnel diagram, equivalent air length. Compare with Fig. 11.17b.

11.7 Gratings

Gratings may be treated in detail with sufficient sampling density to fully resolve the micro structure, see the `grating` command for amplitude gratings and `abr/`(ripple) commands for phase gratings. As it is not practical to use more than a few thousand sample points, the maximum number of grating lines is, in practice, limited to about one thousand or fewer grating lines. For gratings with many thousands of grating lines it is more appropriate to deal only with the macro structure and follow a single grating order. For a given grating order, we may compute the direction of the diffracted light from the Bragg condition, the efficiency considering the profile of the grating line, and aberrations due to the substrate. The `grating/global` command is a nonresolved model of a phase grating.

11.7.1 The Bragg Condition

For a grating having many thousands of grating lines the radiation is sharply defined to the grating orders. Consider a patch of the grating which may be considered effectively flat even for a curved substrate. We have a refracted or reflected wave from the transmission or reflection grating respectively yielding a wave vector $\boldsymbol{k}$ and a diffracted direction wave vector $\boldsymbol{k}'$. $\boldsymbol{\Gamma}$ is the grating vector, m is the diffracted order, and $\hat{\boldsymbol{r}}$ is the direction vector. Let the substrate plane be identified by a normal vector $\hat{\boldsymbol{n}}$. For perfect phase matching across the plane of the substrate we have:

$$\frac{\iint \exp[i(-\boldsymbol{k}' + \boldsymbol{k} + 2\pi\boldsymbol{\Gamma}) \cdot \hat{\boldsymbol{r}}](dx)dy}{\iint dxdy} = 1, \tag{11.116}$$

where the integration is taken over the plane of the substrate. The condition for $\boldsymbol{k}$ to lie in the substrate plane is $\boldsymbol{k} \cdot \hat{\boldsymbol{n}} = 0$. This may be expressed via cross products:

$$\hat{\boldsymbol{n}} \times (\hat{\boldsymbol{k}}' - \hat{\boldsymbol{k}}) = \hat{\boldsymbol{n}} \times m\lambda\boldsymbol{\Gamma}, \tag{11.117}$$

$$\hat{\boldsymbol{n}} \times \hat{\boldsymbol{k}}' = \hat{\boldsymbol{n}} \times (\hat{\boldsymbol{k}} + m\lambda\boldsymbol{\Gamma}). \tag{11.118}$$

We may also write the grating equation in terms of the k-vector projections on the substrate:

$$\hat{\boldsymbol{k}}_{\parallel} = \hat{\boldsymbol{k}} - (\hat{\boldsymbol{k}} \cdot \hat{\mathbf{n}})\hat{\boldsymbol{n}}, \; \hat{\boldsymbol{k}}'_{\parallel} = \hat{\boldsymbol{k}}' - (\hat{\boldsymbol{k}}' \cdot \hat{\mathbf{n}})\hat{\boldsymbol{n}}. \tag{11.119}$$

The grating equation for the in-plane k-vector components is:

$$\hat{\boldsymbol{k}}'_{\parallel} = \hat{\boldsymbol{k}}_{\parallel} + m\lambda\boldsymbol{\Gamma}. \tag{11.120}$$

We may now construct $\hat{\boldsymbol{k}}'$ from $\hat{\boldsymbol{k}}'_{\parallel}$,

$$\hat{\boldsymbol{k}}' = \hat{\boldsymbol{k}}'_{\parallel} + (\hat{\boldsymbol{k}}' \cdot \hat{\mathbf{n}})\hat{\boldsymbol{n}} \tag{11.121}$$

Taking advantage of the fact that $\left|\hat{\boldsymbol{k}}'_{\parallel}\right|^2 + \left|(\hat{\boldsymbol{k}}' \cdot \hat{\mathbf{n}})\right|^2 = 1$, we have

$$\hat{\boldsymbol{k}}' = \hat{\boldsymbol{k}}'_{\parallel} + \frac{\boldsymbol{k} \cdot \hat{\boldsymbol{n}}}{\left|\boldsymbol{k} \cdot \hat{\boldsymbol{n}}\right|} s\sqrt{1 - \left|\hat{\boldsymbol{k}}'_{\parallel}\right|^2}\hat{\boldsymbol{n}} \tag{11.122}$$

$(\boldsymbol{k} \cdot \hat{\boldsymbol{n}})/|\boldsymbol{k} \cdot \hat{\boldsymbol{n}}|$ gives the sign of the incident wave vector relative to the surface normal. s is +1 for a transmission grating and -1 for a reflection grating.

A reflection grating may be viewed by considering the substrate as a mirror followed by a transmission grating. The reflection flips the sign of the components perpendicular to the surface: $|\boldsymbol{k}' \cdot \hat{\boldsymbol{n}}| = -|\boldsymbol{k} \cdot \hat{\boldsymbol{n}}|$ and the Bragg condition from Eq. (11.120) remains the same, affecting only the in-plane components $\boldsymbol{k}_{\parallel}$ and $\boldsymbol{k}'_{\parallel}$. See Fig. 11.19.

An alternate approach to Eq. (11.138) is as follows: define $\boldsymbol{V}$

$$\boldsymbol{V} = \hat{\boldsymbol{k}} + m\lambda\boldsymbol{\Gamma}. \tag{11.123}$$

From Welford the solution is[10]:

$$\hat{\boldsymbol{k}}' = \boldsymbol{V} - (\hat{\boldsymbol{n}} \cdot \boldsymbol{V})\hat{\boldsymbol{n}} + \hat{\boldsymbol{n}}[1 - \boldsymbol{V} \cdot \boldsymbol{V} + (\hat{\boldsymbol{n}} \cdot \boldsymbol{V})^2]^{1/2}. \tag{11.124}$$

We can use $\hat{\boldsymbol{k}}$ and $\hat{\boldsymbol{k}}'$ to find a rotation matrix similar to the method displayed in Eq. (11.53) and Eq. (11.54).

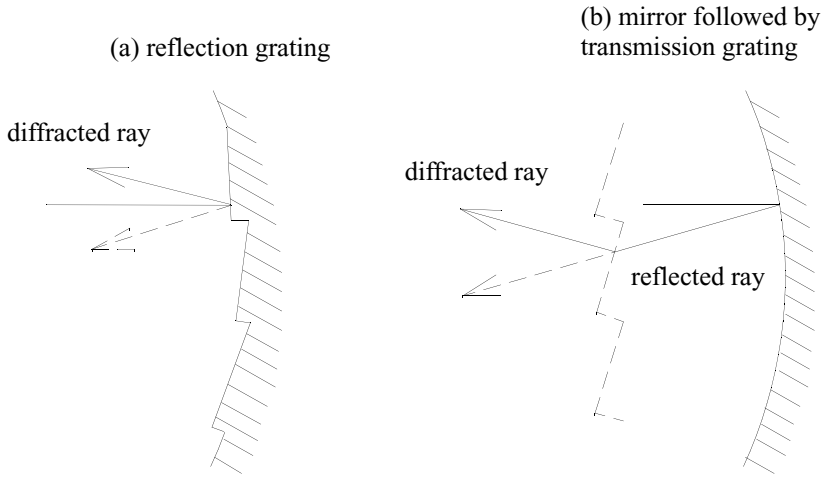


Fig. 11.19. A reflection grating (a) is functionally the same as reflection from the substrate followed by a transmission grating of the same period and orientation (b).

11.8 Gratings (Unresolved Microstructure)

In GLAD, gratings may be treated as resolved or unresolved structures. For the resolved treatment, we must have 6 to 8 sample points over the smallest features. The need to place this many points across each grating line sets a practical limit of a few hundred grating lines to keep the array dimensions below 2048 or 4096. The resolved model has the advantage that all grating orders are calculated simultaneously and the efficiency is calculated from the normal diffraction propagation routines — no special effort is needed.

For gratings with a large number of lines per aperture diameter, the orders may be well separated in angle and it may suffice to model only one diffracted order at a time. In this section we consider how grating efficiency may be calculated. Let us first consider plane gratings and collimated light. Noncollimated beams and curved grating substrates may be treated in local regions by the local grating substrate slope and local beam normals.

Let us consider an incident plane wave of the form

$$\boldsymbol{\psi}_i(\boldsymbol{r}) \;=\; (e_{1_i}\hat{\boldsymbol{i}}_i + e_{2_i}\hat{\boldsymbol{j}}_i)e^{j\boldsymbol{k}_i\cdot\boldsymbol{r}}, \tag{11.125}$$

where e_{1_i} and e_{2_i} are the complex amplitudes of the x- and y-polarization states. Similarly, the exiting beam may be specified by

$$\boldsymbol{\psi}_e(\boldsymbol{r}) \;=\; (e_{1_e}\hat{\boldsymbol{i}}_e + e_{2_e}\hat{\boldsymbol{j}}_e)e^{j\boldsymbol{k}_e\cdot\boldsymbol{r}}, \tag{11.126}$$

where e_{1_e} and e_{2_e} are the complex amplitudes of the x- and y-polarization states of the exiting field. We can also write this as:

$$\boldsymbol{\psi}_e(\boldsymbol{r}) \;=\; e_{1_e}\hat{\boldsymbol{\psi}}_{e_1}(\boldsymbol{r}) + e_{2_e}\hat{\boldsymbol{\psi}}_{e_2}(\boldsymbol{r}), \tag{11.127}$$

$$\hat{\boldsymbol{\psi}}_{e_1}(\boldsymbol{r}) = \hat{\boldsymbol{i}}_e e^{j\boldsymbol{k}_e \cdot \boldsymbol{r}} \text{ and } \hat{\boldsymbol{\psi}}_{e_2}(\boldsymbol{r}) = \hat{\boldsymbol{j}}_e e^{j\boldsymbol{k}_e \cdot \boldsymbol{r}}, \tag{11.128}$$

where $\hat{\boldsymbol{\psi}}_{e_1}(\boldsymbol{r})$ and $\hat{\boldsymbol{\psi}}_{e_2}(\boldsymbol{r})$ are unit vectors for the two orthogonal polarization states of the exiting beam in the $\boldsymbol{k}_e$-direction.

11.8.1 Phase Grating (Unresolved)

The grating will be considered a set of parallel lines lying in a plane defined by $\hat{\boldsymbol{n}}$ with grating frequency vector $\boldsymbol{\Gamma}$. The phase grating has, in the case of a transmission grating, an index of refraction n. In the case of a reflection grating, set $n = -1$. The effect of the grating on the wavefront is:

$$\text{grating function } \phi(x) = \phi_{\text{line}}(x) * |\Gamma| \text{comb}(\Gamma x) e^{j2\pi m \Gamma x}, \tag{11.129}$$

where $\phi(x)$ is the complete wavefront function, x is the coordinate in the plane of the grating perpendicular to the grating lines, $*$ indicates convolution, $\phi_{\text{line}}(x)$ is the wavefront across a single grating period, $\text{comb}(\Gamma x)$ is the sampling function at defining grating lines at interval $1/\Gamma$, and $\exp(j2\pi m \Gamma x)$ is the line-to-line wavefront change.

The coefficients e_{1_e} and e_{2_e} of the exiting beam may be found from the expression for coupled wave efficiency that takes the form of a product of the input mode (Eq. (11.125)), the material equation (grating function, Eq. (11.129)), and the output mode (Eq. (11.128)):

$$e_{1_e} = \frac{\iint_{\text{aperture}} \hat{\boldsymbol{\psi}}^*_{e_1}(\boldsymbol{r}) \cdot \phi(\boldsymbol{r}) \boldsymbol{\psi}_i(\boldsymbol{r}) dx dy}{\iint_{\text{aperture}} dx dy} \text{ and } e_{2_e} = \frac{\iint_{\text{aperture}} \hat{\boldsymbol{\psi}}^*_{e_2}(\boldsymbol{r}) \cdot \phi(\boldsymbol{r}) \boldsymbol{\psi}_i(\boldsymbol{r}) dx dy}{\iint_{\text{aperture}} dx dy}. \tag{11.130}$$

Considering a single grating order m, and taking into account the periodic behavior due to N identical lines:

$$e_{1_e}(m) = (e_{1_i} \hat{\boldsymbol{i}}_e(m) \cdot \hat{\boldsymbol{i}}_i + e_{2_i} \hat{\boldsymbol{i}}_e(m) \cdot \hat{\boldsymbol{i}}_i) \frac{\sum_{n=1}^{N} \left[e^{j(-\boldsymbol{k}_e + \boldsymbol{k}_i + 2\pi\boldsymbol{\Gamma}) \cdot \boldsymbol{r}_n} \int_{\text{line}} \phi_{\text{line}}(\boldsymbol{r}) e^{j(-\boldsymbol{k}_e + \boldsymbol{k}_i) \cdot \boldsymbol{r}} dr \right]}{\sum_{n=1}^{N} \int dr}, \tag{11.131}$$

$$e_{2_e}(m) = (e_{1_i} \hat{\boldsymbol{j}}_e(m) \cdot \hat{\boldsymbol{i}}_i + e_{2_i} \hat{\boldsymbol{j}}_e(m) \cdot \hat{\boldsymbol{i}}_i) \frac{\sum_{n=1}^{N} \left[e^{j(-\boldsymbol{k}_e + \boldsymbol{k}_i + 2\pi m\boldsymbol{\Gamma}) \cdot \boldsymbol{r}_n} \int_{\text{line}} \phi_{\text{line}}(\boldsymbol{r}) e^{j(-\boldsymbol{k}_e + \boldsymbol{k}_i) \cdot \boldsymbol{r}} dr \right]}{\sum_{n=1}^{N} \int dr}.$$

The leading term treats the polarization coupling. The integral evaluates the mode coupling over a single lines under the assumptions that all lines are identical and using the variable **r** for integration that is local to the each line. This method of analyzing the problem is outlined in Fig. 11.20. The global phase matching term is included in the summation over all lines. This exponent of the phasor will sum to zero when *m* takes an integer value. This is the basis of the Bragg condition:.

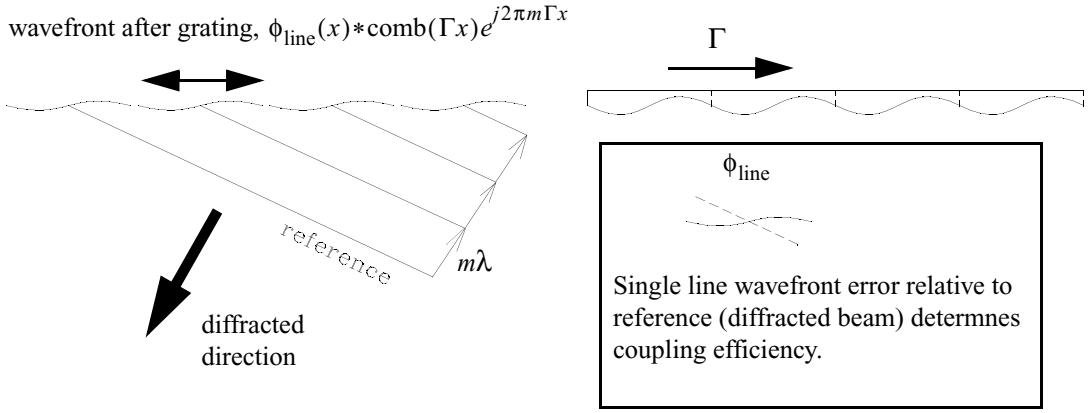


Fig. 11.20. A reflection grating (a) is functionally the same as reflection from the substrate followed by a transmission grating of the same period and orientation (b).

$$\text{Bragg condition } (-\boldsymbol{k}_e(m) + \boldsymbol{k}_i + 2\pi m\boldsymbol{\Gamma}) \cdot \boldsymbol{r} = 0\,. \tag{11.132}$$

When the Bragg Condition is exactly satisfied, all terms in the summation of Eq. (11.131) are identical (within a constant phase term of modulo 2π). We may drop the explicit treatment of the constant modulo 2π phase terms and consider only the integral over a single line and the polarization terms:

$$e_{1_e}(m) = (e_{1_i}\hat{\boldsymbol{i}}_e(m)\cdot\hat{\boldsymbol{i}}_i + e_{2_i}\hat{\boldsymbol{i}}_e(m)\cdot\hat{\boldsymbol{i}}_i)\frac{\int_{\text{line}}\phi_{\text{line}}(\boldsymbol{r})e^{j(-\boldsymbol{k}_e+\boldsymbol{k}_i)\cdot\boldsymbol{r}}\,dr}{\int_{\text{line}}dr}, \tag{11.133}$$

$$e_{2_e}(m) = (e_{1_i}\hat{\boldsymbol{j}}_e(m)\cdot\hat{\boldsymbol{j}}_i + e_{2_i}\hat{\boldsymbol{j}}_e(m)\cdot\hat{\boldsymbol{i}}_i)\frac{\int_{\text{line}}\phi_{\text{line}}(\boldsymbol{r})e^{j(-\boldsymbol{k}_e+\boldsymbol{k}_i)\cdot\boldsymbol{r}}\,dr}{\int_{\text{line}}dr}.$$

The coupling of the input beam into the mth diffracted order is α

$$\alpha(m) = \frac{\int_{\text{line}} \phi_{\text{line}}(\boldsymbol{r}) e^{j(-\boldsymbol{k}_e + \boldsymbol{k}_i)\cdot \boldsymbol{r}} dr}{\int_{\text{line}} dr}, \tag{11.134}$$

$$e_{1_e}(m) = (e_{1_i} \hat{\boldsymbol{i}}_e(m) \cdot \hat{\boldsymbol{i}}_i + e_{2_i} \hat{\boldsymbol{i}}_e(m) \cdot \hat{\boldsymbol{i}}_i)\alpha(m), \tag{11.135}$$

$$e_{2_e}(m) = (e_{1_i} \hat{\boldsymbol{j}}_e(m) \cdot \hat{\boldsymbol{j}}_i + e_{2_i} \hat{\boldsymbol{j}}_e(m) \cdot \hat{\boldsymbol{j}}_i)\alpha(m).$$

We now have the coupling into the exiting diffracted order in terms of the single line coupling efficiency polarization projections.

The coupling term $\alpha(m)$ is related to the Strehl Ratio:

$$\text{Strehl ratio} = |\alpha(m)|^2. \tag{11.136}$$

Evaluation at $m = 0$, yields the ordinary Strehl Ratio associated with the phase modulation of the grating function viewed as aberration. The shape of the grating lines may be optimized such the diffraction efficiency of Eq. (11.134). The term blazing is used when the angle of triangular shaped lines is optimized to direct the input wavefront into the exiting wavefront as efficiently as possible. The optimum angle for blazing may be thought of in term of geometrical optics, applied locally to the grating line, such that an input collimated beam of the proper angle will be refracted or reflected to match the desired output angle that was obtained from the Bragg condition.

The Bragg condition of Eq. (11.132) has an angular sensitivity on the order of $\theta = \lambda / D$ where D is the full width of the grating. Hence the angular sensitivity of the Bragg condition is simply related to the diffraction angle of the full aperture and an error from the Bragg angle causes coupling to drop off as the Strehl ratio of the full aperture drops off with tilt error. The peak coupling at the Bragg angle is determined by the Strehl ratio of the wavefront deformation of a single line. In the case of optimum blazing angle, the coupling factor is $\alpha = 1$and the projection into the polarization components is

$$e_{1_e}(m) = (e_{1_i} \hat{\boldsymbol{i}}_e(m) \cdot \hat{\boldsymbol{i}}_i + e_{2_i} \hat{\boldsymbol{i}}_e(m) \cdot \hat{\boldsymbol{i}}_i), \tag{11.137}$$

$$e_{2_e}(m) = (e_{1_i} \hat{\boldsymbol{j}}_e(m) \cdot \hat{\boldsymbol{i}}_i + e_{2_i} \hat{\boldsymbol{j}}_e(m) \cdot \hat{\boldsymbol{i}}_i).$$

The angular sensitivity to blazing angle is on the order of $\theta = m\lambda / p$ where p is the grating period.

11.8.2 Amplitude Grating (Unresolved)

An amplitude grating has the same Bragg condition and polarization coupling. The coupling efficiency associated with a single line is found from Eq. (11.134) with $\phi(x)$ having the amplitude transmission

function of the absorption grating. The coupling equation Eq. (11.135) may then be used. In the case of a square grating, the amplitude coupling is found in closed-form to be:

$$\alpha = \left(\frac{\sqrt{T+M}-\sqrt{T-M}}{\pi m}\right)\sin\left(\pi m\,\text{Duty}\right)+\left(\frac{\sqrt{T-M}}{\pi m}\right)\sin(\pi m)\,. \tag{11.138}$$

11.9 Decomposing Rotation Matrix into Euler Angles

It is sometimes helpful to interpret the beam matrix **K** in terms of Euler rotation angles. If the beam has passed through an even number of mirrors, then **K** has the properties of a rotation matrix. If the beam has passed through an odd number of mirrors then the parity of **K** is negative so it does not have the properties of a rotation matrix. For this case, the Euler angle decomposition given below will give the direction of the negative k-vector. Consider a rotation matrix

$$\mathbf{R}_{xyz}(\alpha,\beta,\gamma) = \begin{bmatrix} i_x & j_x & k_x \\ i_y & j_y & k_y \\ i_z & j_z & k_z \end{bmatrix} = \begin{bmatrix} R(1,1) & R(1,2) & R(1,3) \\ R(2,1) & R(2,2) & R(2,3) \\ R(3,1) & R(3,2) & R(3,3) \end{bmatrix} = \mathbf{R}_z(\gamma)\mathbf{R}_y(\beta)\mathrm{R}_x(\alpha)\,. \tag{11.139}$$

An arbitrary rotation matrix may be decomposed into Euler angles α, β, and γ as follows. For a beam rotation matrix after an odd number of mirror reflections, the rotation matrix will have odd parity and the k-vector will be flipped.

We proceed to derotate the matrix in the order $\mathrm{R}_z(-\gamma)$, $\mathrm{R}_y(-\beta)$, and $\mathrm{R}_x(-\alpha)$.

Step 1

First rotate by $\mathrm{R}_z(-\gamma)$ by such that the i-vector is parallel to the x-z plane and the y-component is zero

$$\mathbf{R}_z(-\alpha)\begin{bmatrix} i_x \\ i_y \\ i_z \end{bmatrix} = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} i_x \\ i_y \\ i_z \end{bmatrix} = \begin{bmatrix} i'_x \\ 0 \\ i'_z \end{bmatrix} \text{ and } \mathrm{R}_z(-\gamma)\mathbf{R}_{xyz}(\alpha,\beta,\gamma) = \begin{bmatrix} i'_x & j'_x & k'_x \\ 0 & j'_y & k'_y \\ i'_z & j'_z & k'_z \end{bmatrix}. \tag{11.140}$$

We need to solve $-\sin(\gamma)i_x + \cos(\gamma)i_y = 0$.

If $i_x \neq 0$, $\gamma = \tan^{-1}(i_y/i_x)$. There are two solutions: $\gamma_1 = \tan^{-1}(R(2,1)/R(1,1))$ and $\gamma_2 = \gamma_1 + \pi$. We can choose either of two solutions. If $i_x = 0$, then $\gamma_1 = \pi/2$ and $\gamma_2 = -\pi/2$. Again we can choose either of these two solutions. We now implement

$$\mathrm{R}_z(-\gamma)\mathbf{R}_{xyz}(\alpha,\beta,\gamma) = \begin{bmatrix} i'_x & j'_x & k'_x \\ 0 & j'_y & k'_y \\ i'_z & j'_z & k'_z \end{bmatrix}. \tag{11.141}$$

Step 2

We now solve for $R_y(-\beta)$

$$\mathbf{R}_y(-\beta)\begin{bmatrix} i_x' \\ 0 \\ i_z' \end{bmatrix} = \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix}\begin{bmatrix} i_x' \\ 0 \\ i_z' \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \tag{11.142}$$

and

$$R_y(-\beta)R_z(-\gamma)\mathbf{R}_{xyz}(\alpha, \beta, \gamma) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & j_y'' & k_y'' \\ 0 & j_z'' & k_z'' \end{bmatrix}. \tag{11.143}$$

If $i_x' \neq 0$, then $\beta = \tan^{-1}(-i_z'/i_x')$ and the two solutions are $\beta_1 = \tan^{-1}(-R(3,1)'/R(1,1)')$ and $\beta_2 = \beta_1 + \pi$, but only one choice will satisfy the equation $\cos(\beta)i_x' - \sin(\beta)i_z' = 1$. If $i_x' = 0$, we have two solutions $\beta_1 = \pi/2$ and $\beta_2 = -\pi/2$, but only one choice will satisfy the equation $-\sin(\beta)i_z' = 1$.

Step 3

Now rotate by $R_x(-\alpha)$ about the x-axis to align the j-vector with the y-axis,

$$R_x(-\alpha)\begin{bmatrix} 0 \\ j_y'' \\ j_z'' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix}\begin{bmatrix} 0 \\ j_y'' \\ j_z'' \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \tag{11.144}$$

$$R_x(-\alpha)R_y(-\beta)R_z(-\gamma)\mathbf{R}_{xyz}(\alpha, \beta, \gamma) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{11.145}$$

If $j_y'' \neq 0$, then $\alpha = \tan^{-1}(-j_z''/j_y'')$ and the two solutions are $\alpha_1 = \tan^{-1}(-R(3,2)''/R(2,2)'')$ and $\alpha_2 = \alpha_1 + \pi$, but only one choice will satisfy the equation $\cos(\alpha)j_y'' + \sin(\alpha)j_z'' = 1$. If $j_y'' = 0$, we must have $j_z'' = \pm 1$, so we have two solutions $\alpha_1 = \pi/2$ and $\alpha_2 = -\pi/2$, but only one will satisfy the equation $\sin(\alpha)j_z'' = 1$.

References

1. Menchaca and D. Malacara, "Directional curvatures in a conic surface", Appl. Opt. 23, 3258 (1 Oct 1984).

2. W. Brouwer, *Matrix Methods in Optical Instrument Design*, W. A. Benjamin, Inc., 2465 Broadway, New York NY 10025 (1964).

3. L. Levi, *Applied Optics*, John Wiley & Sons, New York (1968).
4. R. E. Hopkins, “Mirror and Prism Systems”, *Applied Optics and Optical Engineering*, Vol III, Ed. R. Kingslake, Academic Press.
5. T. Smith, Trans. Opt. Soc. (London) 30, 68 (1927-1928).
6. John C. Polasek, “Matrix Analysis of Gimbaled Mirror and Prism Systems”, JOSA Vol 57, No. 10, (1967).
7. David A. Berkowitz, “Design of Plane Mirror Systems”, JOSA Vol. 55, No. 11 (1965).
8. Joseph S. Beggs, “Mirror-Image Kinematics”, JOSA Vol. 50, No. 4 (1960).
9. MIL-HDBK-141, U. S. Military Standardization Handbook for Optical Design, US Government Printing Office, Sections on Geometrical Optics and Optical Design available from Sinclair Optics.
10. W. T. Welford, *Aberrations of Optical Systems*, Section 5.4 and 5.5, Adam Hilger, (1986).

12. Atmospheric Propagation

This chapter discusses many of the features necessary to model the propagation of laser beams over long distances. The major considerations are atmospheric turbulence, adaptive optics, jitter of the beam, thermal blooming and integration of the irradiance pattern on a distant target. Each of these effects is relatively well understood in its own right, but the possible interactions are complex and numerical calculations of the full system performance are frequently required to gain an understanding of the system performance. Figure 12.1 shows a representative system which might be used for laser communication. A laser beam is generated on the ground. The beam is projected by a beam expander to a relay optic in space. In the uplink path, the beam is affected by aberrations in the beam expander, the adaptive optic element, and atmospheric optics. The adaptive optic is controlled to correct the aberrations observed in a downward path through the atmosphere form a beacon placed on (or close to) the relay mirror platform. The relay optic intercepts the beam and projects it to a distant focusing mirror. The finite relay mirror clips the energy distribution from the ground-based laser. The relay mirror aperture acts to some extent as a spatial filter so that there is relatively little preservation of the aberrations from the beam expander and uplink atmospheric effects. These aberrations have their principal effect in reducing the energy transmission. The relayed beam is projected to the focusing mirror by propagation through space.

The focusing mirror has aberrations due to intrinsic shape deformation and due to the effects of rapid steering. If there is a long distance from the relay to the focusing mirror there will be a spatial filtering effect similar to that of the uplink path to the relay mirror. It is generally most cost-effective to let the side lobes of the diffraction pattern fall outside the aperture of the focusing mirror rather than require a larger aperture for either the relay mirror or the focusing mirror. Because of spatial filtering effects of the uplink and relay-to-focusing propagations and the mirror apertures, only the aberrations of the focusing mirror directly effect the irradiance distribution at the target. Atmospheric aberrations in the downlink path have relatively little effect on the target irradiance because the aberrations are added close to the target, essentially in the near field of the target.

12.1 Atmospheric Aberration

The atmospheric turbulence model is based on an angular spectrum model which is readily implemented and which is readily interfaced with the adaptive optics model. There is a lack of agreement among the atmospheric experts with regard to proper treatment of atmospheric outer scale. This leads to uncertainty in atmospheric tilt and piston error which is problematic for uncorrected systems and for large phased arrays but not for adaptively corrected systems. Atmospheric aberration is described by the Kolmogorov spectral distribution. The aberration is described by an integral over the altitude in the atmosphere[1].

$$W^2(\rho) = \frac{0.38}{\lambda^2 \rho^{11/3}} \int_{h_{\min}}^{h_{\max}} C_n^2(h)\,dh\,, \tag{12.1}$$

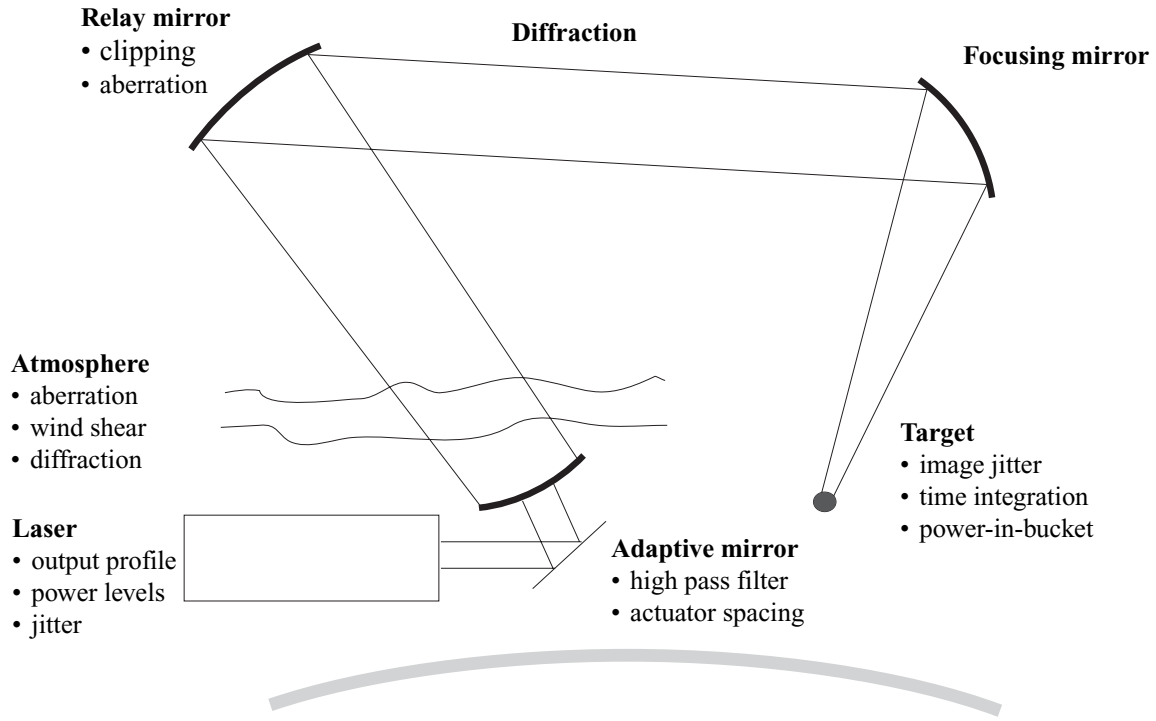


Fig. 12.1. Representative laser communication system showing the ground-based laser, the aberrations of the atmosphere, propagation to the relay mirror, aperturing by the relay mirror, propagation to the focusing mirror, and downlink propagation to a target on the ground.

where $C_n^2(h)$ is the refractive index structure constant of the atmosphere as a function of altitude, h is the altitude in meters, ρ is the spatial frequency variable, λ is the wavelength, and $W^2(\rho)$ is the power spectrum of the wavefront. We may characterize the aberration by the autocorrelation diameter, r_0, sometimes known as Fried's parameter

$$r_0 = \left[0.423 k^2 \int_{h_{\min}}^{h_{\max}} C_n^2(h) dh \right]^{-5/3} . \tag{12.2}$$

Fried's parameter may then be used to calculate the wavefront

$$W^2(\rho) = \frac{0.23}{r_0^{5/3} \rho^{11/3}} . \tag{12.3}$$

The inner and outer scale may be included according to Lutomirski and Yura [2]

$$W^2(\rho) = \frac{0.023 e^{-\rho^2 L_i^2}}{r_0^{5/3}\left(\rho^2 + \frac{1}{L_o^2}\right)^{11/6}}, \tag{12.4}$$

where ρ is the spatial frequency, L_o is the outer scale, and L_i is the inner scale. These parameters are in radians, meters, and inverse meters respectively. L_o is generally taken to be several meters and L_i to be on the order of centimeters or less. From the relationships described above, we can see that it is possible to characterize the atmospheric aberration by Fried's parameter. It is natural therefore to separate the problem of atmospheric characterization into three parts: 1) determination of the effects of various levels of atmosphere in terms of r_0, 2) summation of the r_0's, and 3) implementation of aberration using the r_0 parameter. The summation of the r_0's of different levels takes the form

$$r_{\text{total}} = [r_1^{-5/3} + r_2^{-5/3} + \ldots]^{-3/5}. \tag{12.5}$$

Propagation through N layers of equivalent aberration of r_0, which might occur in horizontal propagation, will result in

$$r_{\text{total}} = \frac{r_0}{N^{3/5}}. \tag{12.6}$$

Propagation over considerable distances is accomplished by taking short steps of simple diffraction alternating with short steps of aberration introduction. Figure 12.2 illustrates schematically the procedure.

The number of steps depends on the magnitude and spatial frequencies of the aberration. $C_n^2(h)$ varies with altitude. An approximate expression which is adequate for estimation purposes is

$$C_n^2(h) = 2\times10^{-12} h^{-4/3}. \tag{12.7}$$

Day time performance is about an order of magnitude lower. Consideration of Eq. (12.7) will reveal that the most important factor is the height of the exit pupil of the telescope which sets the lower limit of the integral. For a wavelength of 0.5×10^{-6} meters and the simplified expression for $C_n^2(h)$, Eq. (12.2) can be evaluated. For a lower limit of $h_1 = 10$ meters, $r_0 \approx 3$ cm. For day time conditions, $r_0 \approx 0.7$ cm. The seeing may be significantly improved by raising the height of the laser source and by designing the laser station to minimize local turbulence. The $C_n^2(h)$ calculations are difficult and have significant uncertainties. For this reason it is more convenient to develop our aberration expressions in terms of r_0 and leave the calculation of that parameter to the user to determine by experiment or by calculation.

An average height may be calculated by Eq. 12.8

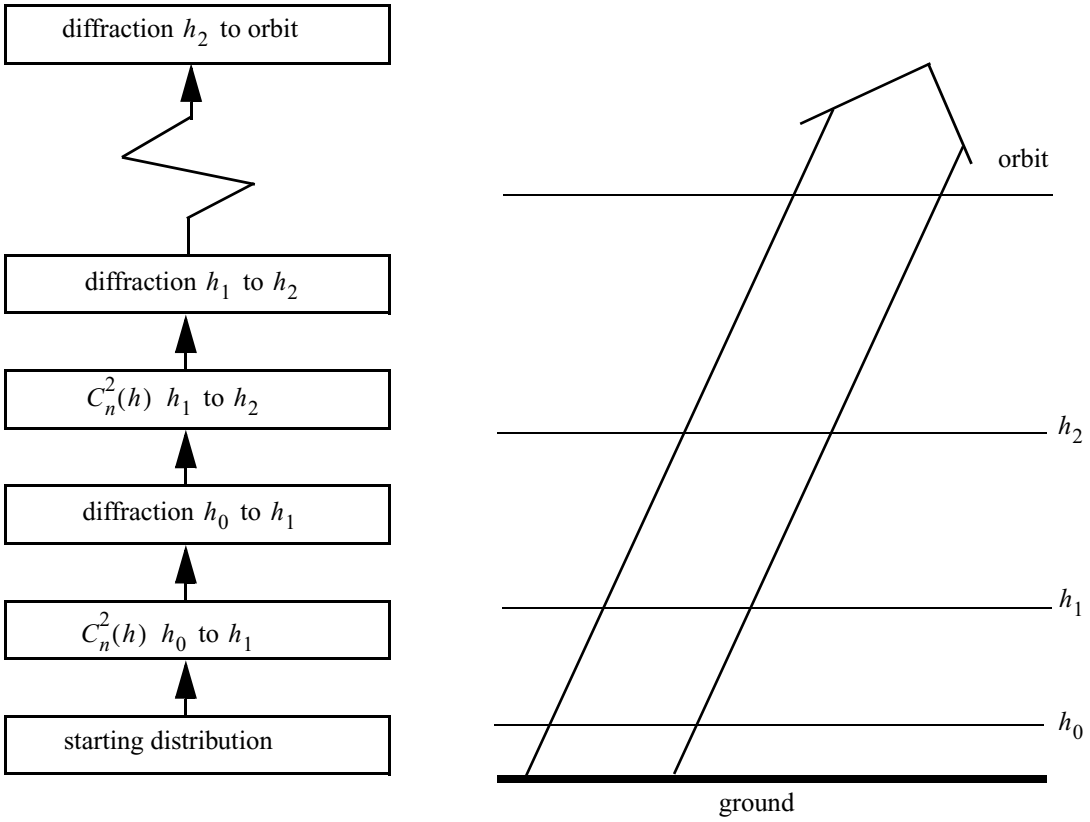


Fig. 12.2. Schematic of propagation through the air where aberration and diffraction propagation must be considered. The propagation is done in relatively short steps with the diffraction propagation and addition of aberration alternated to achieve an approximation to a continuous process.

$$\langle h \rangle = \frac{\int_{10}^{10^4} h h^{-4/3} dh}{\int_{10}^{10^4} h^{-4/3} dh}. \tag{12.8}$$

This equation depends upon the upper limit of the atmosphere which is not well determined and which really represents the point at which Eq. 12.7 ceases to be valid and the aberration falls off to zero. However, Eq. 12.8 indicates that the most of the aberration occurs at relatively low altitude. The Strehl ratio is found from the approximate expression

$$\mathrm{SR} = e^{-\int W^2(\omega) d\rho}. \tag{12.9}$$

The frequency spectrum falls off rapidly and we can identify a maximum such that frequencies above the maximum need not be considered. Assuming a propagation distance of about 500 meters, the seeing parameter is about r_0 = 5 cm. Frequencies above 48 m^{-1} contribute only a 2 percent drop in Strehl Ratio. The diffraction characteristic length is

$$z_{\text{char}} = \frac{1}{2\lambda\rho_{\max}^2} \approx 430 \text{ meters.} \tag{12.10}$$

This would indicate the use of several levels: perhaps 10-500, 500-1,500, and 1,500-10,000 meters. The aberration due to the atmosphere is a smoothed random wavefront. However it changes over time because of beam steering and wind shear. This effect may be readily modeled by using Fourier transform methods to shift the distribution. Given a wavefront of $w(x, y)$ and velocity components v_x and v_y, the shift in the atmospheric aberration with time is

$$w(x - v_x t, y - v_y t) = \text{FF}^{-1}[e^{-2\pi(v_x\xi + v_y\eta)t} FF[w(x, y)]], \tag{12.11}$$

where FF indicates a two-dimensional Fourier transform. Figure 12.3 shows a typical atmospheric aberration phase sheet with the wind blowing in the +x direction. Because of the cyclical nature of FFT methods the aberration that disappears off one side reappears on the other.

12.2 Adaptive Optics

Because of the high spatial frequencies present in the atmospheric aberration, any aperture of significant size requires a large number of actuators ranging from several hundred to several thousand. Methods developed for treating adaptive correction of thermal mirror distortion which employed tens of actuators and which used individually specified influence functions for each actuator are neither practical nor necessary. A zonal approach which uses an angular spectrum technique allows rapid calculation of the adaptive optic correction. The method is shown to be traceable to the optimum least squares fitting approach and may be used for steady-state influence functions or transient influence functions [Ref. 3–6,8–9].

Let p_i be points in the array and $m_j(x)$ be the influence function for the j^{th} actuator. The mirror surface may be described as the linear superposition of the influence functions

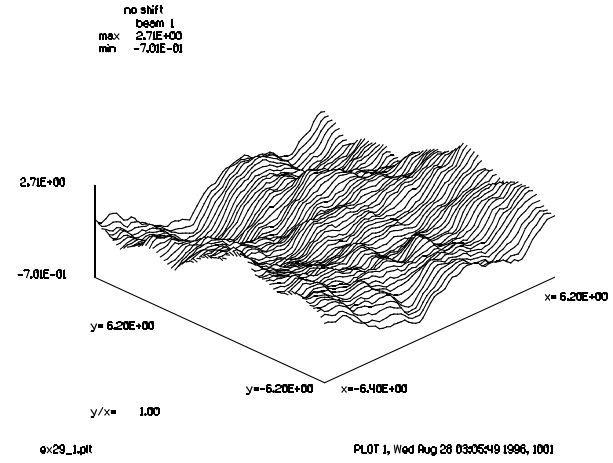
$$w_m(p_i) = \sum_j a_j m_j(p_i). \tag{12.12}$$

This may be expressed in matrix and vector form

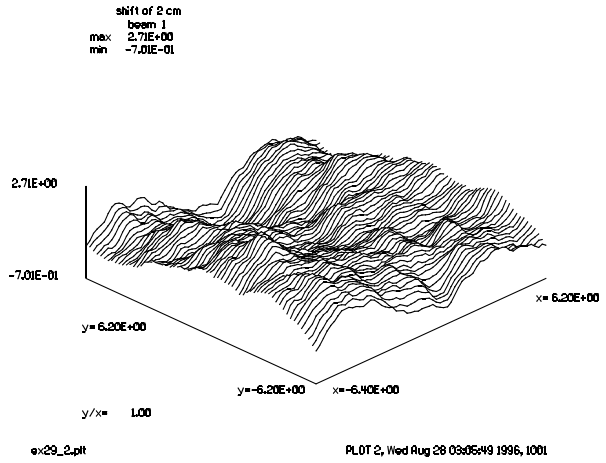
$$\boldsymbol{w}_m = \mathbf{H}\boldsymbol{a}, \tag{12.13}$$

where $\boldsymbol{w}_m$ and $\boldsymbol{a}$ are the wavefront and actuator coefficient vectors and the matrix. **H** is defined such that

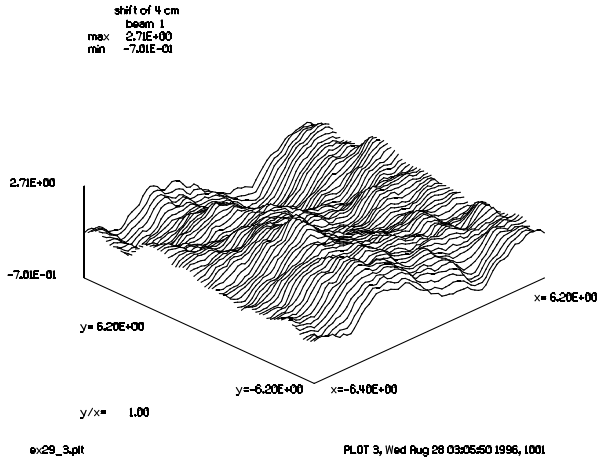
$$h_{ij} = m_j(p_i). \tag{12.14}$$



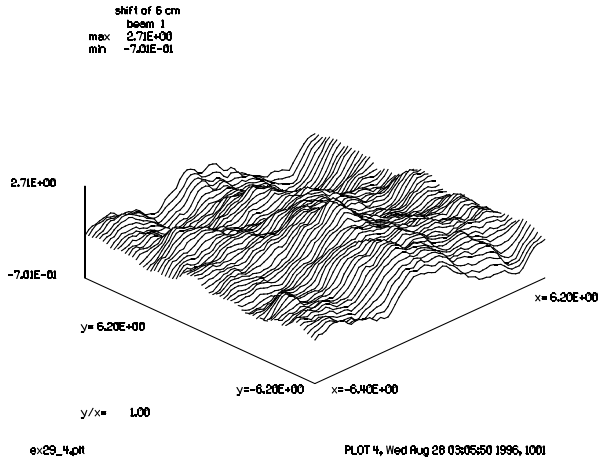
(a) No shift.



(b) Shift of 2 cm.



(c) Shift of 4 cm.



(d) Shift of 6 cm.

Fig. 12.3. Atmospheric aberration wavefront based on the Kolmogorov spectrum with a wind blowing in the direction. It can be seen that the aberration which disappears off the +x edge reappears on the -x edge.

Let the wavefront we intend to fit be $w_m(p_i)$. We attempt to fit the mirror to the wavefront for minimum error

$$\varepsilon(p_i) \;=\; w(p_i) - w_m(p_i)\,. \tag{12.15}$$

The wavefront variance is

$$\sigma^2 \;=\; \boldsymbol{\varepsilon}^t\boldsymbol{\varepsilon}\,. \tag{12.16}$$

The solution for minimum variance is the well known least squares solution [6],

$$\boldsymbol{a} \;=\; (\mathbf{H}^t\mathbf{H})^{-1}\mathbf{H}^t\boldsymbol{w}, \tag{12.17}$$

where $\boldsymbol{w}$ is the wavefront to be fit. The best-fit wavefront is

$$\boldsymbol{w}_m = \mathbf{H}(\mathbf{H}^t\mathbf{H})^{-1}\mathbf{H}^t\boldsymbol{w}, \tag{12.18}$$

$$\boldsymbol{w}_m = \mathbf{S}\boldsymbol{w} \text{ where } \mathbf{S} = \mathbf{H}(\mathbf{H}^t\mathbf{H})^{-1}\mathbf{H}^t. \tag{12.19}$$

For a typical case of 100 actuators and a square computer array of 64×64 points, direct evaluation of Eq. (12.19) would require tens of millions of multiplications.

A particularly simple solution is possible when it is assumed that all actuators are identical. In this case

$$m_j(p_i) = m(x_i - x_{c_j}, y_i - y_{c_j}), \tag{12.20}$$

where $(p_i = (x_i, y_i))$ is the point to be calculated and (x_{c_j}, y_{c_j}) is the actuator location which is the center of the influence function. The influence function we should use is the influence obtained by operation of the adaptive mirror with the optimum control law. The influence function obtained by using the optimum control law is the narrowest possible influence function. Mathematically, this is equivalent to orthogonalization of the influence factors. The real influence functions are, in general, not orthogonal. However, we can construct an orthonormal set from the original set. Let the orthonormal set have influence function matrix $\mathbf{H}'$ and coefficient vector $\boldsymbol{a}'$, such that

$$\boldsymbol{w}_m = \mathbf{Ha} = \mathbf{H}'\boldsymbol{a}'. \tag{12.21}$$

If we have the orthonormalized coefficients $\boldsymbol{a}'$ we can determine the actual coefficients for the actuators by

$$\boldsymbol{a} = \mathbf{H}^{-1}\mathbf{H}'. \tag{12.22}$$

In terms of the orthonormal matrix $\mathbf{H}'$ the system matrix $\mathbf{S}$ is

$$\mathbf{S} = \mathbf{H}(\mathbf{H}^t\mathbf{H})^{-1}\mathbf{H}^t = \mathbf{H}'\mathbf{H}'^t. \tag{12.23}$$

This operation reflects the operation of the optimum actuator control law. We can find the orthonormal set by the Grahm-Schmidt process. The Grahm-Schmidt process [7] assumes the function set to be ordered. The j^{th} orthonormalized influence function is

$$m_j' = \frac{m_j - \sum_{k=0}^{j-1} \langle m_j m_k' \rangle m_k'}{\left| m_j - \sum_{k=0}^{j-1} \langle m_j m_k' \rangle m_k' \right|}. \tag{12.24}$$

Figure 12.4 represents graphically the form of **H′** and **S**. Figure 12.5 shows an isometric plot of a typical **S** matrix formed from orthonormalized influence functions..

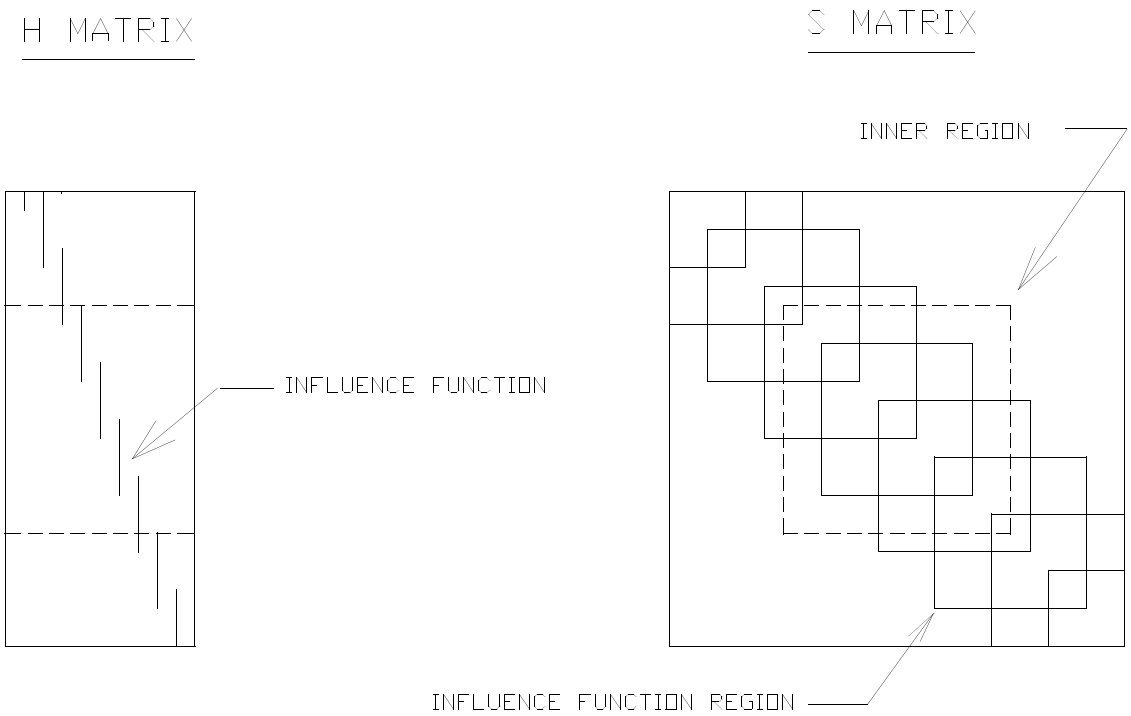


Fig. 12.4. Form of **H′** and **S**. **H′** is a rectangular matrix with each column representing the orthogonal influence function of one of the actuators. Except near the edges the columns are identical except shifted. **S**, the system matrix, has the form indicated on the right.

For an infinite array of actuator influence functions on a rectangular grid, the orthonormalization process is modified. Since all influence functions are assumed to be identical, we need orthogonalize only one influence function against its neighbors. Generally we will need to orthogonalize only to the nearest neighbors or, at most, to the second nearest neighbors unless the mirror face plate is so stiff that the influence function spans many actuators. Mathematically the orthonormalization process can be represented as

$$m' = \frac{m - \sum_{j=1}^{N}\sum_{k=1}^{N} \langle mm(\pm j\Delta x, \pm k\Delta y)\rangle m(\pm j\Delta x, \pm k\Delta y)}{\left| m - \sum_{j=1}^{N}\sum_{k=1}^{N} \langle mm(\pm j\Delta x, \pm k\Delta y)\rangle m(\pm j\Delta x, \pm k\Delta y) \right|}, \tag{12.25}$$

where Δx and Δy are the actuator spacings and N is the order of the neighbors considered. The ± sign indicates that we are to include positive and negative values of the function shift in the summation. To include only nearest neighbors, $N = 1$.

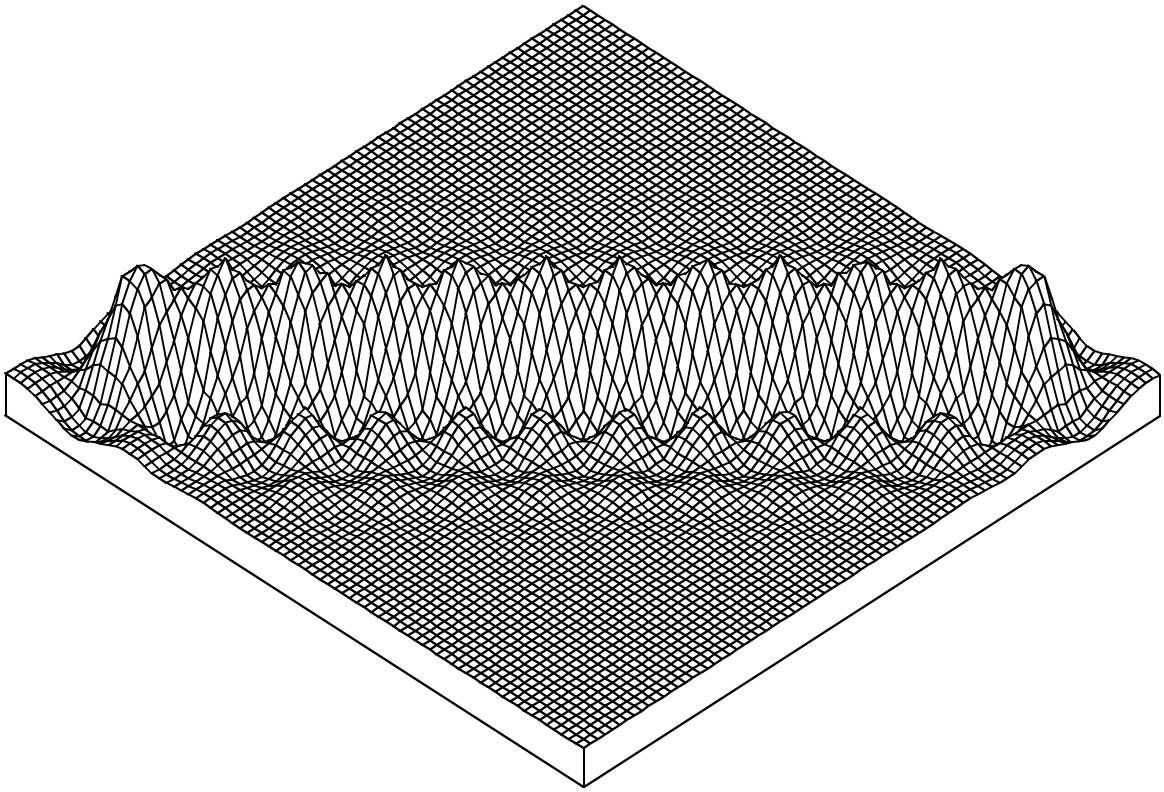


Fig. 12.5. Isometric of typical **S** matrix formed from orthonormalized influence functions.

Equation (12.25) shows that the neighboring points are activated to remove the mirror movement to the extent possible. The orthogonal function represents the condition where the actuator control law has been used to minimize the width of the influence function for one actuator by manipulating the neighboring actuators. The orthogonal influence function, therefore, represents the influence function of the actuator under optimum control. Figure 12.6 shows schematically the uncontrolled and controlled, orthogonal influence functions.



Fig. 12.6. The uncontrolled influence function (left) is found by moving one actuator and keeping all other actuators fixed (assuming position actuators). The controlled influence function (right) is somewhat narrower because the neighboring influence functions are moved to oppose the effect of the central actuator. The controlled influence function is orthogonal to its neighbors.

In the remainder of this discussion, we shall assume that the influence functions have been orthonormalized and drop the primes from the nomenclature for $m(p_i)$. Given *m* orthonormalized then the appropriate equation is

$$\boldsymbol{w}_m = \mathbf{H}\mathbf{H}^t\boldsymbol{w}, \tag{12.26}$$

$$\boldsymbol{\varepsilon} = \mathrm{I} - \mathbf{H}\mathbf{H}^t\boldsymbol{w}. \tag{12.27}$$

Consider the case of relatively fine sampling of the wavefront such that there are many sample points between actuators. In which case $\mathbf{H}$ is a high aspect ratio rectangular array with many more rows than columns. Since $\mathbf{H}^t$ consists of a series of identical influence functions which are shifted to have their centers at the locations of the actuators, we have the relationship based on the comb function

$$\boldsymbol{a} = \mathbf{H}^t \rightarrow \sum_i a(p_i)\delta(p_i) = (m(x,y)**w(x,y))\mathrm{comb}\left(\frac{x}{\Delta x}, \frac{y}{\Delta y}\right), \tag{12.28}$$

where actuator differences near the boundaries are ignored. The above relationship derives from the fact that all influence functions are identical. This point may be more clear by considering a one-dimensional example. Eq. 12.29 illustrates the appearance of the matrix-vector operations for one-dimension. The coefficients of the actuator are the dot products of the wavefront with the shifted influence functions.

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \cdot \end{bmatrix} = \begin{bmatrix} m_1 & m_2 & m_3 & 0 & 0 & 0 & 0 & 0 & 0 & \cdot \\ 0 & 0 & 0 & m_1 & m_2 & m_3 & 0 & 0 & 0 & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & m_1 & m_2 & m_3 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \\ w_7 \\ w_8 \\ w_9 \\ \cdot \end{bmatrix}. \tag{12.29}$$

Similarly

$$\boldsymbol{w}_m = \mathbf{H}\boldsymbol{a} \rightarrow w_m \approx m(x,y)**\sum_i a_i\delta(p_i), \tag{12.30}$$

where $\delta(p_i)$ is a delta function at point p_i.

The net result of $\mathbf{H}\mathbf{H}^t$ is

$$w_m(x,y) = m(x,y)**\left\{[m(x,y)**w(x,y)]\mathrm{comb}\left(\frac{x}{\Delta x}, \frac{y}{\Delta y}\right)\right\}. \tag{12.31}$$

In frequency space this is

$$W_m(\xi, \eta) = M(\xi, \pi)[M(\xi, \eta)W(\xi, \eta)**|\Delta x||\Delta y|\text{comb}(\Delta x\xi, \Delta y\eta)]\,. \tag{12.32}$$

In frequency space the error function is

$$\text{E}(\xi, \eta) = 1 - M(\xi, \pi)[M(\xi, \eta)W(\xi, \eta)**|\Delta x||\Delta y|\text{comb}(\Delta x\xi, \Delta y\eta)]\,. \tag{12.33}$$

If the wavefront is sampled by a detector of finite area $d(x, y)$ then Eq. (12.31) is modified to be

$$w_m(x, y) = m(x, y)**\left\{[m(x, y)**(w(x, y)**d(x, y))]\text{comb}\left(\frac{x}{\Delta x}, \frac{y}{\Delta y}\right)\right\} \tag{12.34}$$

and the error function becomes

$$\text{E}(\xi, \eta) = 1 - M(\xi, \pi)[M(\xi, \eta)W(\xi, \eta)D(\xi, \eta)**|\Delta x||\Delta y|\text{comb}(\Delta x\xi, \Delta y\eta)]\,. \tag{12.35}$$

The convolution in frequency space result in aliasing which represents printthrough of the actuators — a significant effect if the actuator spacing is too large for the mirror stiffness.

Figure 12.7 illustrates the spatial and frequency space aspects of the adaptive mirror process. The original wavefront is illustrated in the top of the figure. In general we can expect a mixture of high and low spatial frequency errors in the wavefront. The influence function is illustrated as a gaussian-like function. Because of orthogonalization the function dips below the axis to remove the effects of the adjacent influence functions. Convolution of the wavefront by $m(x, y)$ smooths out the high spatial frequencies.

The actuators exist at discrete points separated by Δx and Δy. The coefficients are represented by delta functions of varying height. In the spatial domain, the smoothed wavefront is multiplied by the appropriate comb function. In frequency space, the damped frequency function is convolved by the Fourier transformed comb function, which is itself a comb function. The convolution of the comb function may be considered as a series of the damped frequency functions. These functions will, in general, overlap resulting in aliasing. The overlap is the result of the influence functions being too narrow or their spacing being too wide. This will occur if the faceplate of the active mirror is not sufficiently stiff. The effect is sometimes referred to as actuator print through. To make maximum use of a given number of actuators it will be best to design the faceplate stiffness to allow some print through. The actuator coefficients multiply the influence functions centered at the actuator locations. This may be represented by a convolution of the actuator delta functions by the influence functions. Figures 12.8a–12.8d show examples of orthogonalized influences functions in the spatial and frequency domains for various coupling percentages between neighboring actuators: 5%, 15%, and 50%.

In frequency space this is a multiplication again by the Fourier transform of the influence function. However, some of the aliased information is left in the frequency pass band. The residual errors are represented by the difference between the adaptive mirror and the original wavefront. The residual wavefront consists principally of the high spatial frequencies (the part of the original wavefront which falls

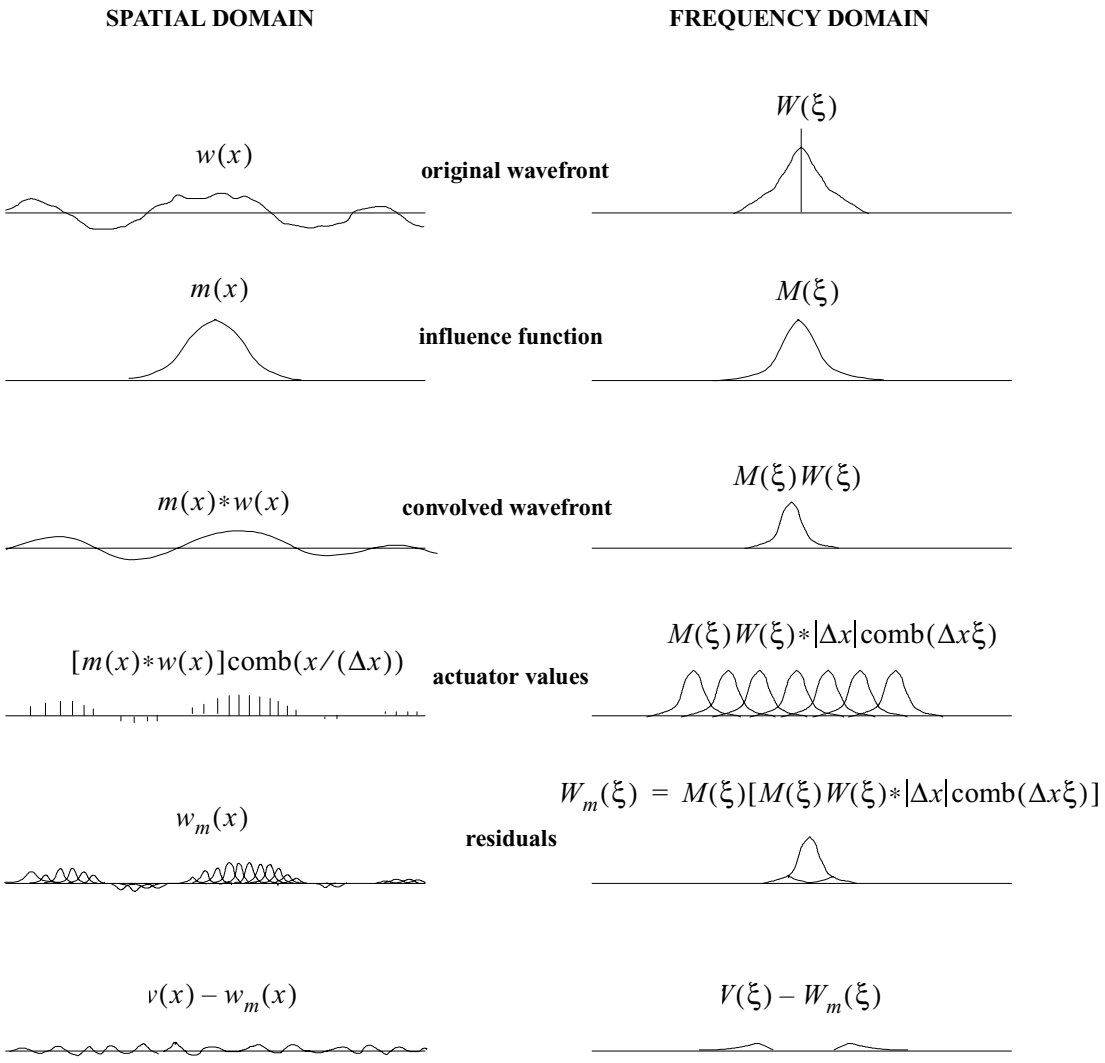


Fig. 12.7. Spatial and frequency space representations of the functional steps in the adaptive optic modeling. The original wavefront and influence functions are shown. Convolution by the influence function and subsequent sampling at the actuator spacing gives the coefficients of the influence functions. The coefficients establish the size of the influence functions at the actuator points. The frequency space representation shows aliasing in the frequency domain which translates into printthrough in the spatial domain. The residual wavefront shows the effects of high spatial frequencies which can not be corrected by the actuators and printthrough at the period of the actuator spacing.

outside the spectral band of the influence function) and the print through effects and the period of the actuators.

The model described above will provide a good approximation to the full least squares influence function fitting model provided the actuators fall upon a rectangular grid. The principal errors are on the perimeter of the mirror where the actuator performance is likely to be affected by the boundary of the mirror and, therefore, will not satisfy our condition of identical actuators. Alda and Boreman have extended this model to include edge effects [9]. For a stiff faceplate and an ideal control system, the controlled influence function is

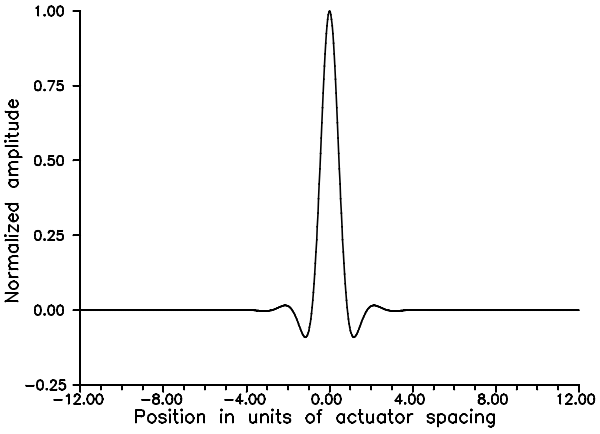


Fig. 12.8a. Influence function with 5% coupling.

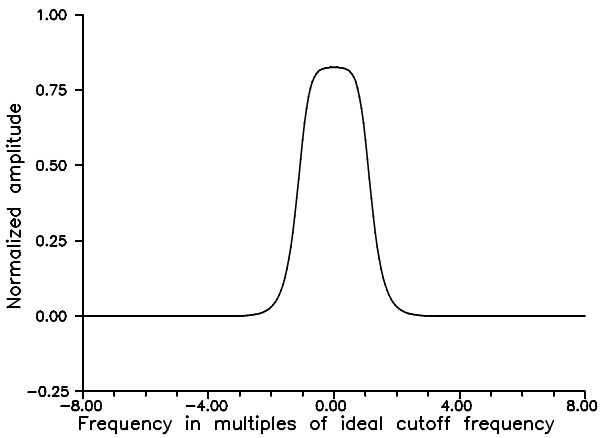


Fig. 12.8b. Fourier transform of influence function with 5% coupling.

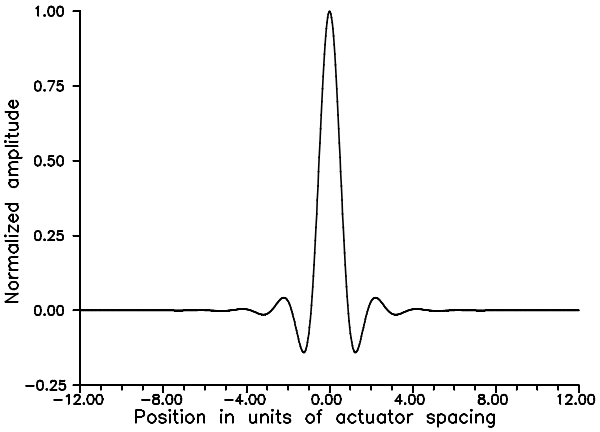


Fig. 12.8c. Influence function with 15% coupling.

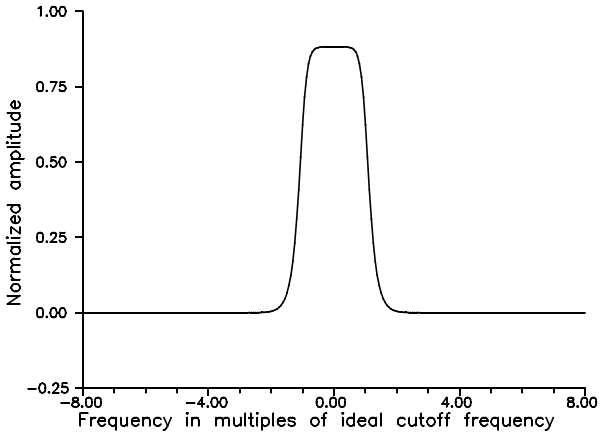


Fig. 12.8d. Fourier transform of influence function with 15% coupling.

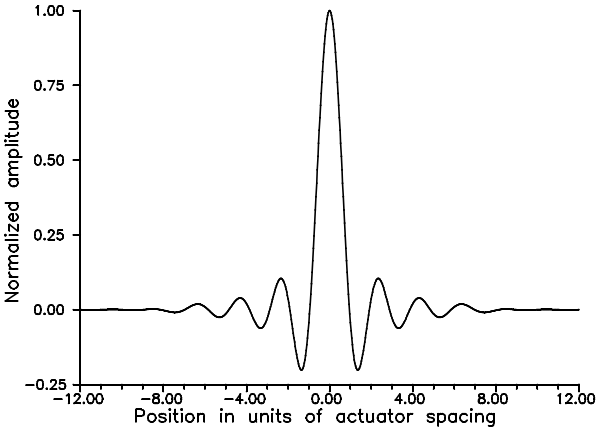


Fig. 12.8e. Influence function with 50% coupling.

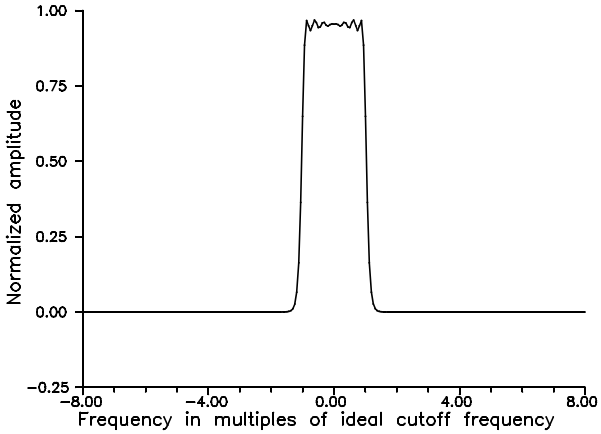


Fig. 12.8f. Fourier transform of influence function with 50% coupling. Nearly a rect function.

$$m(x) = \frac{\sin\left(\frac{\pi x}{2\Delta x}\right)}{\frac{\pi x}{2\Delta x}}. \tag{12.36}$$

Figures 12.9.a shows an example of aberration correction by an adaptive mirror. Figure 12.9.a shows a wavefront with atmospheric aberration. Figure 12.9b shows the residual wavefront after low order aberrations are removed by the adaptive optic. High spatial frequencies remain which are of relatively low amplitude.

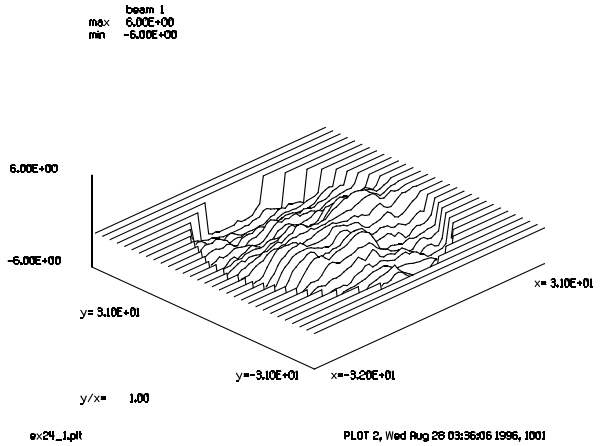


Fig. 12.9.a. Wavefront before correction.

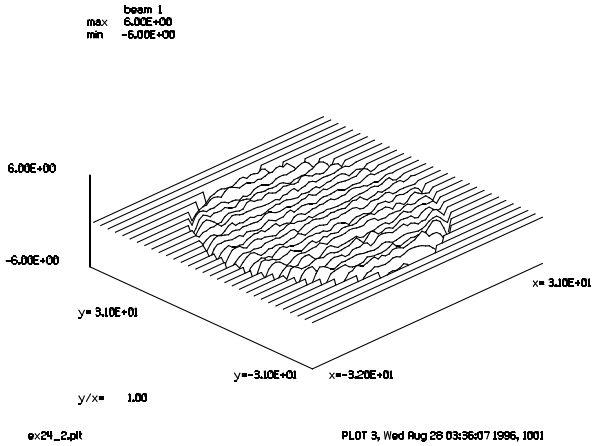


Fig. 12.9b. Wavefront after correction.

References

1. Roddier, “The Effects of Atmospheric Turbulence in Optical Astronomy,” E. Wolfe, Ed., *Progress in Optics XIX*, North-Holland (1981).
2. Lutomirski and H. T. Yura, “Aperture Averaging Factor for a Fluctuating Light Signal,” J. Opt. Sac. Am., Vol. 59, No. 9, pp 1247-1248 (1969)
3. James E. Harvey and Gary M. Callahan, “Wavefront Error Compensation Capabilities of Multi-Actuator Deformable Mirrors,” SPIE Proc. 141, 50 (1978).
4. R. K. Tyson and D. M. Byrne, SPIE Proc. 228, 21 (1980).
5. Robert K. Tyson, “Using the deformable mirror as a spatial filter: application to circular beams,” Appl. Opt. Vol. 21, No. 5, 787 (1982).
6. Charles L. Lawson and Richard J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ (1974).
7. G. Arfken, *Mathematical Methods for Physicists*, Academic Press, New York, NY, 2nd Ed. (1970).
8. K. E. Moore and G. N. Lawrence, “Zonal Model of an Adaptive Mirror,” Appl. Opt, Vol. 29, pp 4622-4628 (1990).
9. J. Alda and G. D. Boreman, “Analysis of edge effects for deformable mirrors,” Opt. Eng. Vol. 31, pp 2282-2286 (1992).

13. Thermal Blooming

In this section, the effects of atmospheric absorption of a high power beam and the resulting laser heating (or cooling) of the atmosphere are described. As the beam heats the atmosphere the index of refraction is altered. The net effect is to change the direction and the degree of convergence of the beam. Much of the original work was done in the US in the 1970's and was rather completely documented in the open literature at that time and it is from this body of work that the techniques described here were developed. References [1–5] fully describe the scientific and technological basis of the thermal blooming model. Reference [1] by Charles B. Hogge and Ref. [2] by Strohbehn provides a comprehensive description of thermal blooming as applied to high power CO_2 lasers. Reference [3] by Wood, Camac and Gerry is a good source of optical properties of the atmosphere. Reference [4] is a thorough discussion of thermal blooming and addresses the particular problem of imaging through the perturbed atmosphere. Reference [4] includes a complete listing of HELPRP which incorporates the thermal blooming models of the other HELP codes. References [5] and [6] are standard US references on properties of the atmosphere. Reference [7] is a convenient computer program based on Reference [6] from which some parts of the GLAD atmospheric routines were adapted. Reference 8 is a standard reference for the physical and optical properties of air.

The basic effects to be included are listed in Table 13.1. Given the altitude, the code will calculate the temperature, pressure, index of refraction, and absorption of the atmosphere. The absorption of nitrogen and carbon dioxide are broken out separately to enable the effects of kinetic cooling to be included. Kinetic cooling is a transient phenomena in which, if the conditions are right, the gas is initially cooled by CO_2 absorption of light and then heated as the stimulated states of CO_2 spread to N_2 and then decay thermally. The effect may be important if the beam is crossing through the air very rapidly. The effect may reduce or reverse the thermal defocusing of the atmosphere. The user may also explicitly define the atmospheric parameters rather than use the values calculated by GLAD. The attenuation effects of the atmosphere are included as well as the phase perturbations due to temperature change in the air. The relative rate at which the beam crosses through the air due to wind and beam steering may be specified. The phase change causes the beam to change its degree of convergence, direction, and amount of aberration. The code will test for excessive growth of the beam within the bounds of the computer array and automatically rescale and recenter the beam as required.

Table. 13.1. Thermal blooming features.

- automatic calculation of the state of the atmosphere vs. altitude
- user explicit specification of atmospheric parameters
- absorption due to water and carbon dioxide
- phase perturbations due to kinetic heating and cooling
- beam shear effects from wind and beam steering
- automatic resizing and centering of the beam in the computer array

13.1 Mathematical Derivations and Techniques

In order to calculate the aberrations of an optical beam induced in the atmosphere by self heating, it is necessary to solve two major problems: 1) the change in index of the air due to absorbed power and 2) the transverse profile of the temperature changes. One of the key aspects of the problem is the beam crossing time. The beam crossing time, τ_{trans}, is the time for the atmosphere to sweep past the beam due to the combined effects of beam steering and wind. In performing these calculations we shall make the following assumptions:

- The beam has been on for at least one beam crossing time,
- The beam crossing time is finite. Completely stagnant air is not modeled,
- Transient phenomena have time constants less than the time for sound to cross the beam,
- The speed of sound is approximately 330 m/sec at standard temperature and pressure (STP),
- The temperature distribution is "frozen" into the air.,No movement of heat is considered.
- Only 10.6 μ effects are considered,
- The beam crossing rate is constant over one beam crossing time.

The thermal blooming effects will become infinitely large if completely stagnant air is modeled because of factors of $1/v$ that appear in the equations derived below. In practice heat would dissipate out of the beam area by convection and conduction but the beam would be so disrupted by the extreme heating that the beam quality would degenerate to the point of being useless. Because the approximations used here do not allow for conduction or convection, the heat is considered to be frozen in place where it is deposited by the laser beam. This frozen distribution passes across the beam because of the differential rates of beam steering and wind as laser power is continually absorbed. The deposited energy is smeared into the atmosphere in much the same way that a pen puts ink down on paper.

The heating of the atmosphere causes the gas to expand at constant pressure. The expansion takes place at the rate of the sound propagation in the air. As the volume increases the density drops proportionally. The equation for index change is given by Eq. 13.1 from Reference [1]

$$\Delta n = -79\times10^{-6}\frac{P}{T^2}\Delta T, \tag{13.1}$$

where P is in millibars and T is in degrees Kelvin. The equation may also be written

$$\Delta n = -(n_0 - 1)\frac{P_{rel}}{T_{rel}}\frac{\Delta T}{T}, \tag{13.2}$$

where P_{rel} is the relative atmospheric pressure and T_{rel} is the relative temperature with respect to standard temperature and pressure (STP) and n_0 is the index of refraction at 10.6 μ and STP, n_0 = 1.000274[6]

$$\phi = -\frac{2\pi}{\lambda}(n_0 - 1)\frac{P_{rel}}{T_{rel}}\frac{\Delta T}{T}\Delta z, \tag{13.3}$$

n_0	n_0 is the index of refraction air at STP,
P_{rel}	Relative pressure,

T_{rel} Relative temperature with respect to 288 K,
ΔT Temperature rise due to absorption,
T Temperature of the atmosphere,
Δz Propagation step length,
λ Wavelength.

The next step is to determine the temperature rise. The attenuation of the beam is characterized by the loss coefficient α km^{-1} which includes all losses. For 10.6 μ radiation the principal loss is due to absorption by water vapor, carbon dioxide, and other components of the atmosphere. It is assumed that scattering plays a negligible role and that all of the attenuation is due to energy absorption. The computation of the temperature rise is the most difficult part of the problem. In simple form, the equation for temperature to be solved is

$$\Delta T = \frac{\alpha}{\rho C} \int_{-\infty}^{0} I(\tau) d\tau , \tag{13.4}$$

where

α is the attenuation constant [m^{-1}],
ρ is the density [kg/m^3],
$I(\tau)$ is the irradiance [w/m^2],
C is the absolute specific heat [J/kg/deg],
τ is the time.

The integration limits are $-\infty$ to 0, but the irradiance $I(\tau)$ is zero after approximately one beam crossing time. This point is clear from the description in terms of time and position is

$$I(x, y, \tau) = I(x - \Delta x(\tau), y - \Delta y(\tau)) , \tag{13.5}$$

where the spatial shifts $\Delta x(\tau)$ and $\Delta y(\tau)$ are defined by

$$\Delta x(\tau) = \left[v_{w_x} + \frac{s}{s_t} v_{t_x} \right] \tau = v_{c_x} \tau , \tag{13.6a}$$

$$\Delta y(\tau) = \left[v_{w_y} + \frac{s}{s_t} v_{t_y} \right] \tau = v_{c_y} \tau , \tag{13.6b}$$

v_{c_x}, v_{c_y} net components of velocity,
v_{w_x}, v_{w_y} components of wind velocity,
v_{t_x}, v_{t_y} components of target velocity,
s slant range of the current beam position,
s_t slant range of the target.

For sufficiently short intervals of Δz along the optical path, the slant range does not vary significantly and the velocity components may be considered constant over that interval. As shown above, the temperature is based on the time integration of the beam moving in the atmosphere. We may disregard the

times sufficiently far back that no part of the beam overlaps the position of interest. A practical limitation is that the edge of the computer array defines the most distant time to be considered. Consequently the time integration may begin at approximately one beam crossing time in the past. For a computer array of sides N and sample spacing of Δx the crossing time is approximately

$$\tau_{\text{trans}} \approx \frac{N\Delta x}{v}, \tag{13.7}$$

where v is the beam crossing velocity. For reference, some of the properties of air are listed in Table 13.2.

Table. 13.2. Physical and optical constants for air.

index of refraction (10.6 μ)	2.74×10^{-4}
density [kg/m^3]	1.226
conductivity [w/mK]	0.02525
specific heat [J/kgK]	1017

Under proper conditions, a CO_2 laser beam may cause kinetic cooling. This occurs as a transient condition when the CO_2 absorption in the atmosphere is significant. According to Reference [1], as the CO_2 molecule absorbs radiation in going from the (1000) state to the (0001) state, energy is absorbed from the vibrational (1000) level. Consequently (1000) is no longer in thermal equilibrium and the state is repopulated quickly by taking energy out of the translational energy reservoir. The process results in an immediate drop in temperature. Ultimately the energy in the (0001) state distributes its energy into the first vibrational state of nitrogen. The excited nitrogen atoms decay slowly through any of the available routes, so that for some modest period of time the effect of the laser is to cool the gas. Since a typical value for the nitrogen decay time constant is $\tau_{N_2} \approx 0.015$ sec., the kinetic cooling may be a factor if the beam crossing time is less than the nitrogen decay time

$$\frac{\tau}{\tau_{N_2}} \ll 1. \tag{13.8}$$

This condition may occur for fast beam steering resulting in a cancellation or reversal of the usual kinetic heating effect. Kinetic cooling is incorporated by Hogge into the equations as shown below. Let α_{H_2O} and α_{CO_2} be the absorption coefficients of water vapor and carbon dioxide respectively. Then the total absorption is

$$\alpha = \alpha_{H_2O} + \alpha_{CO_2} + \ldots, \tag{13.9}$$

$$\sigma = 2.441(n_0 - 1)\frac{\alpha_{CO_2}}{\alpha}. \tag{13.10}$$

The full equation including kinetic cooling is

$$\Delta T(x,y) = -\frac{\alpha}{\rho C_p}\int_{-\infty}^{0}\left(1-\sigma e^{-\frac{\tau}{\tau_{N_2}}}\right)I(x-v_{c_x}\tau, y-v_{c_y}\tau)d\tau, \tag{13.11}$$

α total atmospheric coefficient = $\alpha_{H_2O} + \alpha_{CO_2}$,
ρ density of air 1.226 kg/m3,
C_p specific heat 1017 J/kgK,
τ time variable,
τ_{N_2} time constant of nitrogen for kinetic cooling,
σ kinetic cooling factor, Eq. (13.10).

The integral over time may be accomplished by performing a moving average over the irradiance distribution. For general components this requires an interpolation step across matrix lines which is numerically noisy. When kinetic cooling is negligible, a particularly smooth way to do this is to perform the moving average by doing the integration in the frequency domain.

13.2 Blooming Algorithms

There are three basic forms in which the calculation may be performed. We shall use two constants in each case:

$$C = -\frac{2\pi}{\lambda}(n_0-1)\frac{\alpha}{\rho C T}\Delta z \tag{13.12}$$

and σ from Eq. (13.10)

1) Convert from complex amplitude to intensity and phase

$$a(x,y) \rightarrow I(x,y)e^{jW(x,y)}. \tag{13.13}$$

2) Calculate wavefront change

$$W(x,y) = \int_{\tau=0}^{\tau_{\min}} C\left(1-e^{-\frac{\tau}{\tau_{N_2}}}\right)I(x-v_{c_x}\tau, y-v_{c_y}\tau)d\tau. \tag{13.14}$$

3) Convert from intensity and phase to complex amplitude

$$I(x,y)e^{jW(x,y)} \rightarrow a(x,y). \tag{13.15}$$

This is the default method and is considered the most reliable.

`bloom/prop/xonly`

If the crossing velocity is strictly along the x-direction considerable speed improvements are possible. The thermal blooming and kinetic cooling are based on two sums taken over the array. The code scans in

the x-direction then the y-direction because data is stored contiguously along the rows. The intermediate sums are:

$$\text{Intensity sum: Sum1} = \sum_{k=-\infty}^{0} I(x - k\Delta x, y), \tag{13.16}$$

$$\text{kinetic cooling sum with temporal decay: Sum2} = \sum_{k=-\infty}^{0} e^{-\frac{k\Delta\tau}{\tau_{N_2}}} I(x - k\Delta x, y), \tag{13.17}$$

represented as a series:

$$\text{Sum2} = e^{-\frac{\Delta\tau}{\tau_{N_2}}}\left(I(x,y) + e^{-\frac{\Delta\tau}{\tau_{N_2}}}\left(I(x-\Delta x, y) + e^{-\frac{\Delta\tau}{\tau_{N_2}}} I(x - 2\Delta x, y) + \ldots\right)\right). \tag{13.18}$$

The exponentials may be eliminated by using an intermediate value

$$r_{\text{fac}} = e^{-\frac{\Delta\tau}{\tau_{N_2}}}, \tag{13.19}$$

$$\text{Sum2} = r_{\text{fac}}(I(x,y) + r_{\text{fac}}(I(x-\Delta x, y) + r_{\text{fac}}(I(x - 2\Delta x, y) + \ldots))), \tag{13.20}$$

The computer coding takes the form

$$\text{Sum2} = r_{\text{fac}}[I(x) + \text{Sum2}], \tag{13.21}$$

$$\phi = C\Delta\tau(\text{Sum1} - \sigma\text{Sum2}). \tag{13.22}$$

`bloom/prop/fft`

The temporal integration may be performed in the frequency domain. This method avoids the spatial domain shifting which is rather time consuming. The Fourier method is derived below:

$$\int_{\tau_{\min}}^{0} \left(1 - \sigma e^{-\frac{\tau}{\tau_{N_2}}}\right) I(x - v_x\tau, y - v_y\tau) d\tau, \tag{13.23}$$

$$= \mathrm{FF}^{-1}\left[\int_{\tau_{\min}}^{0}\left(1-\sigma e^{-\frac{\tau}{\tau_{N_2}}}\right)\mathrm{FF}[I(x,y)]e^{j2\pi(\xi v_x\tau+\eta v_y\tau)}d\tau\right], \tag{13.24}$$

$$= \mathrm{FF}^{-1}\left[\int_{\tau_{\min}}^{0}\left(1-\sigma e^{-\frac{\tau}{\tau_{N_2}}}\right)I(\xi,\eta)e^{j2\pi(\xi v_x\tau+\eta v_y\tau)}d\tau\right], \tag{13.25}$$

$$= \mathrm{FF}^{-1}\left[\int_{\tau_{\min}}^{0}\left[I(\xi,\eta)e^{j2\pi(\xi v_x\tau+\eta v_y\tau)}-\sigma I(\xi,\eta)e^{-\frac{\tau}{\tau_{N_2}}+j2\pi(\xi v_x\tau+\eta v_y\tau)}\right]d\tau\right], \tag{13.26}$$

$$= \mathrm{FF}^{-1}\left[\mathrm{FF}[I(x,y)]\left[\frac{1-e^{j2\pi(\xi v_x+\eta v_y)\tau_{\min}}}{j2\pi(\xi v_x+\eta v_y)}-\frac{\sigma\left(1-e^{-\frac{\tau}{\tau_{N_2}}+j2\pi(\xi v_x+\eta v_y)\tau_{\min}}\right)}{-\frac{\tau}{\tau_{N_2}}+j2\pi(\xi v_x+\eta v_y)}\right]\right]. \tag{13.27}$$

We choose $\tau_{\min}$ to be the time that results in the beam being 1/4 of the array width from the center. Choosing a larger value will result in "wrapping" of the intensity distribution. There is still some slight wrapping with the current algorithm. Wrapping effects are minimized by windowing the intensity distribution to the inner one-half rectangle. The phase calculations are followed by an internal call to `fitphas` to remove tilt and focus error and to invoke a diffraction propagation. If `Nstep` is specified then successive steps of kinetics and diffraction will be used — each of length `Z/Nstep`. The tilt and focus terms are removed by fitting the aberration polynomials to the wavefront. The first step is to convert from complex amplitude to intensity and phase according to Eq. (13.13)

Form the inner products of the wavefront with the four lowest order Zernike polynomials and of the four lowest order polynomials with each other

$$w_c = w(x,y) - \sum_{i=1}^{4} a_i z_i . \tag{13.28}$$

The equation to be solved to determine the coefficients a_i is

$$\begin{bmatrix} \langle z_1 w \rangle \\ \langle z_2 w \rangle \\ \langle z_3 w \rangle \\ \langle z_4 w \rangle \end{bmatrix} = \begin{bmatrix} \langle z_1 z_1 \rangle & \langle z_1 z_2 \rangle & \langle z_1 z_3 \rangle & \langle z_1 z_4 \rangle \\ \langle z_2 z_1 \rangle & \langle z_2 z_2 \rangle & \langle z_2 z_3 \rangle & \langle z_2 z_4 \rangle \\ \langle z_3 z_1 \rangle & \langle z_3 z_2 \rangle & \langle z_3 z_3 \rangle & \langle z_3 z_4 \rangle \\ \langle z_4 z_1 \rangle & \langle z_4 z_2 \rangle & \langle z_4 z_3 \rangle & \langle z_4 z_4 \rangle \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}, \tag{13.29}$$

where the inner product is defined to be

$$\langle \psi_1 \psi_2 \rangle = \iint I(x, y) \psi_1(x, y) \psi_2(x, y) dx dy, \tag{13.30}$$

so that the inner product is a function of the irradiance distribution. These functions will, in general, not be orthogonal over the particular distribution. This equation is solved and the low order terms are removed. The intensity and phase are then converted back to complex amplitude by Eq. (13.15).

It is interesting to observe that the development above does not require the use of gain sheets or of explicit time treatment. This greatly speeds the computation and simplifies the programming. Figure 13.1a shows the phase distribution due to thermal heating with wind blowing in the -y direction. The heated air stretches from the center to the -y edge. The phase distribution is shown after being fitted for tilt and focus. Since the fit of Eq. (13.30) is over a gaussian distribution in the center of the array, the tilt and focus are best-fit over that area and show significant tilt and focus error outside that region. Fig. 13.1b show the characteristic "sugar scoop" effect with the formation of a satellite hot spot down stream from the wind.

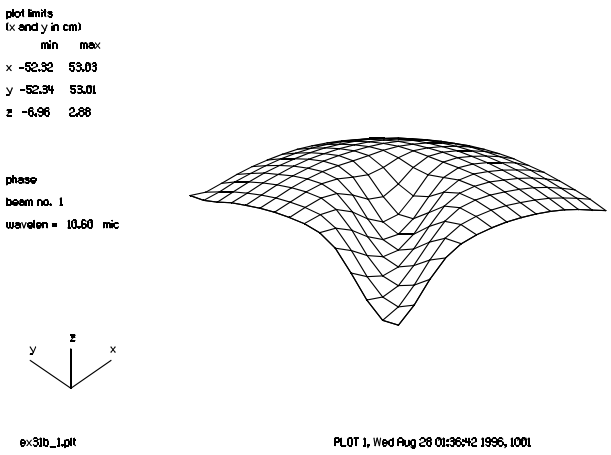


Fig. 13.1a. Phase after a step of thermal blooming. The wind is blowing in the -y direction, causing a trough in the phase. The heated air acts to diverge the light. The tilt and focus have been best-fit over the center of the array where the gaussian beam intensity is significant, causing the phase outside the beam to show tilt and focus error.

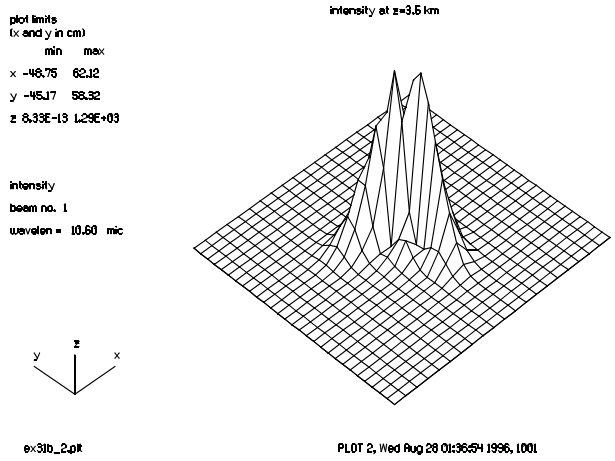


Fig. 13.1b. Intensity distribution after propagation with thermal blooming. The initial distribution was gaussian. The "sugar scoop" is a characteristic feature as is the secondary peak which forms down stream.

References

1. Charles B. Hogge, "Propagation of High Energy Laser Beams in the Atmosphere," in *High Energy Lasers and Their Applications*, S. Jacobs, et. al., Eds., Addison-Wesley, Reading Massachusetts (1974).
2. J. L. Walsh and P. B. Ulrich, "Thermal Blooming in the Atmosphere," *Laser Beam Propagation in the Atmosphere*, Ed. J. W. Strohbehn, Springer-Verlag, New York (1978).
3. A. D. Wood, M. Camac and E. T. Gerry, Appl. Opt. 10, 1877 (1971).
4. David A. Nahrstedt, "The Influence of a Thermally Bloomed Atmosphere on Target Image Quality," PhD Dissertation, University of Arizona (1981).
5. U. S. Air Force, Handbook of Geophysics, revised edition, McMillan and Co., New York (1960).
6. U. S. Standard Atmosphere 1976, National Oceanographic and Atmospheric Administration, Washington DC, 1976.
7. Frank J. Regan, "An Atmosphere Model," ACCESS, LEDS Publishing Co., Nov/Dec 1987.
8. Handbook of Chemistry and Physics, Chemical Rubber Company, E204, 52 Ed. (1971).

14. Polarization

The state of polarization plays an important role in determining the performance of optical systems. A uniform polarization state across the pupil is a requirement for achieving good beam quality and for optimum coherent combination of beams. Many optical components alter the polarization state of optical beams. In a complex optical system, with three-dimensional folds and tilted components, the polarization state can be altered significantly as it passes through the system. Most importantly, the different areas of the aperture may be affected differently if the incidence angles are different, which lowers the beam quality and focusability of the beam. The GLAD code, in its present form, treats light as being strictly monochromatic. Monochromatic light, is of necessity, perfectly polarized. Partial polarization requires a finite spectral bandwidth with varying polarization states in the different spectral components. For full polarization, the polarization state at any point in the aperture may be represented by 2-vectors and the transformations by 2×2 matrices, rather than the more complex 4-vectors and 4×4 Mueller matrices required for partial polarization. By convention, the field at a given point in space is

$$\boldsymbol{E}(x, y) = E_x(x, y)\hat{\boldsymbol{x}} + E_y(x, y)\hat{\boldsymbol{y}}, \tag{14.1}$$

where $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are unit vectors in the x- and y-directions. In keeping with the use of beam coordinates $\hat{\boldsymbol{i}}$, $\hat{\boldsymbol{j}}$, and $\hat{\boldsymbol{k}}$, which move with the beam. We shall use

$$\boldsymbol{E}(x, y) = E_x(x, y)\hat{\boldsymbol{i}} + E_y(x, y)\hat{\boldsymbol{j}}, \tag{14.2}$$

where (x, y) are calculated along the local unit vectors $\hat{\boldsymbol{i}}$ and $\hat{\boldsymbol{j}}$. This allows the beam to rotate as a unit without changing the apparent state of polarization. In column vector form, the optical field is

$$\boldsymbol{E}(x, y) = \begin{bmatrix} E_x(x, y) \\ E_y(x, y) \end{bmatrix}. \tag{14.3}$$

We show several typical Jones vectors in Table 14.1. The sense of circular polarization is right handed if clockwise when viewed looking back down the direction of propagation. The polarization state may be modified by a large variety of components, but all possible polarization alterations effects may be characterized by a 2×2 matrix, called the Jones matrix

$$\mathbf{J} = \begin{bmatrix} j_{11} & j_{12} \\ j_{21} & j_{22} \end{bmatrix}, \tag{14.4}$$

Table. 14.1. Common polarization states.

$\begin{bmatrix}1\\0\end{bmatrix}$	x linear polarization
$\begin{bmatrix}0\\1\end{bmatrix}$	y linear polarization
$\frac{\sqrt{2}}{2}\begin{bmatrix}1\\1\end{bmatrix}$	+45º linear polarization
$\frac{\sqrt{2}}{2}\begin{bmatrix}1\\1\end{bmatrix}$	–45º linear polarization
$\frac{\sqrt{2}}{2}\begin{bmatrix}1\\-j\end{bmatrix}$	right circular polarization
$\frac{\sqrt{2}}{2}\begin{bmatrix}1\\j\end{bmatrix}$	left circular polarization

where the elements are complex coefficients. The 4 complex elements account for 8 degrees of freedom. Following Hecht[2], we show several typical Jones matrices in Table 14.2.

14.1 Rotation of Polarization Components

The Jones matrices may be rotated to represent the effect of a rotated polarization active component by

$$\mathbf{J}_{\text{rotated}} = \mathbf{R}(\theta)\mathbf{J}\mathbf{R}^{-1}(\theta), \tag{14.5}$$

where

$$\mathbf{R}(\theta) = \begin{bmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix}. \tag{14.6}$$

A clockwise rotation is assumed to be positive. Using the rotation matrices, we can define any polarizer and retarder in terms of a rotated vertical polarizer and a rotated vertical fast axis retarder

$$\text{rotated vertical polarizer} = \mathbf{R}(\theta)\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\mathbf{R}^{-1}(\theta), \tag{14.7}$$

Table. 14.2. Typical Jones matrices.

$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	x linear polarizer
$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$	y linear polarizer
$\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$	linear polarizer at +45º
$\frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$	linear polarizer at –45º
$e^{j\frac{\pi}{4}}\begin{bmatrix} 1 & 0 \\ 0 & -j \end{bmatrix}$	quarter-wave plate, vertical fast axis
$e^{j\frac{\pi}{4}}\begin{bmatrix} 1 & 0 \\ 0 & j \end{bmatrix}$	quarter-wave plate, horizontal fast axis
$\frac{1}{2}\begin{bmatrix} 1 & j \\ -j & 1 \end{bmatrix}$	right circular polarizer
$\frac{1}{2}\begin{bmatrix} 1 & -j \\ j & 1 \end{bmatrix}$	left circular polarizer

$$\text{rotated vertical fast axis retarder} = \mathbf{R}(\theta)\begin{bmatrix} e^{j\frac{\Psi}{2}} & 0 \\ 0 & e^{-j\frac{\Psi}{2}} \end{bmatrix}\mathbf{R}^{-1}(\theta)\,. \tag{14.8}$$

The right and left circular polarizers are not affected by rotation.

14.1.1 Distributed Effects

For polarization effects that take place in a distributed manner during propagation through a material, the Jones matrix procedure may be modified to take the form

$$\frac{\partial E}{\partial z} = \mathbf{J}E\,, \tag{14.9}$$

where the elements of **J** have units of inverse distance. Equation (14.9) may be integrated to give

$$E(\Delta z) = e^{\mathbf{J}\Delta z}. \tag{14.10}$$

The exponential of the Jones matrix is an operator, which may be evaluated by a series expansion of matrix elements. Consider a matrix **A**; the exponential is

$$e^{\mathbf{A}} = \sum_{k=0}^{\infty} \frac{\mathrm{A}^k}{k!}. \tag{14.11}$$

14.2 Common Polarization Effects and Components

In this section, many common types of components are considered which can alter the polarization state.

14.2.1 Birefringence

Many crystals are anisotropic, i.e., they have different properties in different directions. Such crystals have different indices for different polarization states and different directions of propagation. The current treatment in the GLAD code is limited to uniaxial with the crystal axis perpendicular to the propagation direction. In this case, we can consider a fast axis and slow axis for the crystal. It is common to consider an ordinary and extraordinary direction and associated indices n_o and n_e, such that $n_o > n_e$. Since light travels faster in the lower index, the fast axis corresponds to the extraordinary ray.

In the more general treatment of crystal axis orientation, the ordinary and extraordinary rays propagate in different directions. This induces a shearing in the beams which is not, at present, considered in the GLAD program. With the optical axis perpendicular to the direction of propagation, the birefringent material acts as a retarder. The equation for retardance is

$$\Delta\psi = \frac{2\pi}{\lambda} d(n_o - n_e). \tag{14.12}$$

With the fast axis vertical, the Jones matrix is given by Eq. 14.8.

14.2.2 Reflecting Mirror

The beam matrix is transformed by a plane mirror such that (see Eq. (11.45a)),

$$\left[\hat{\boldsymbol{i}}'\ \hat{\boldsymbol{j}}'\ \hat{\boldsymbol{k}}'\right] = \left[\hat{\boldsymbol{i}}\ \hat{\boldsymbol{j}}\ \hat{\boldsymbol{k}}\right] - 2\left[(\hat{\boldsymbol{i}}\cdot\hat{\boldsymbol{n}})\hat{\boldsymbol{n}}(\hat{\boldsymbol{j}}\cdot\hat{\boldsymbol{n}})\hat{\boldsymbol{n}}(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{n}})\hat{\boldsymbol{n}}\right] \tag{14.13}$$

This operation flips the parity of the beam. The Jones matrix for a reflector is simply,

$$\text{Jones matrix for reflection} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{14.14}$$

The complex amplitudes E_x and E_y remain unchanged but the direction of the $\hat{\boldsymbol{k}}$ changes with respect to $\hat{\boldsymbol{i}}$ and $\hat{\boldsymbol{j}}$. This reverses the definition of left and right circular polarization. The parity of the beam is defined by (See also Eq. (11.82)).

$$\hat{\boldsymbol{k}} \cdot (\hat{\boldsymbol{i}} \times \hat{\boldsymbol{j}}) . \tag{14.15}$$

If Eq. (14.13) is positive the parity is right-handed and left-handed if negative.

14.2.3 Optical Activity

Optical components which cause a rotation of the orientation of linear polarization are called optically active. Crystalline quartz, sugar solutions, tartaric acid, and turpentine exhibit optical activity. If the rotation is clockwise it is termed d-rotatory and termed l-rotatory if counterclockwise. This is due to different retardance of left and right circular polarization. The angle of rotation, relative to the direction of propagation, is determined from the equation[1],

$$\theta_k = \frac{\pi}{\lambda} |\boldsymbol{d} \cdot \hat{\boldsymbol{z}}| (n_L - n_R) , \tag{14.16}$$

$\boldsymbol{d}$	propagating direction vector (cm),
$\hat{\boldsymbol{z}}$	axis of crystal,
n_L	index of left circular light,
n_R	index of right circular light.

The Jones matrix for optical activity is

$$\mathbf{R}(\theta_k) = \begin{bmatrix} \cos(\theta_k) & \sin(\theta_k) \\ -\sin(\theta_k) & \cos(\theta_k) \end{bmatrix} . \tag{14.17}$$

With optical activity the polarization rotates with respect to the direction of propagation. Propagation through the material followed by reverse propagation after reflection from a mirror will result in cancellation of the effects because the angle θ_k is taken with respect to the direction of optical propagation, which is reversed in a backward pass. This is comparable to the rotation associated with advancing along a mechanical screw. A forward advance along a right handed screw followed by backward propagation would result in zero net rotation.

14.2.4 Faraday Effect

Faraday discovered that application of an external magnetic field could cause a rotation of the orientation of linear polarization. The governing equation for the Faraday Effect is

$$\theta_k = V(\boldsymbol{B} \cdot \boldsymbol{d}) , \tag{14.18}$$

$\boldsymbol{d}$	propagation direction vector (cm),

V Verdet constant,

$\boldsymbol{B}$ magnetic flux vector density (gauss).

A positive sign of the Verdet constant is, by convention, l-rotary. The Faraday effect is similar to optical activity except that the angle accumulates in the backward propagation direction. The direction of rotation of the phase always follows the direction of rotation of current in a coil that would generate the $\boldsymbol{B}$ field. The angular direction of rotation reverses with a change in propagation direction. If a mirror is used to reverse the direction of the beam, the handedness of the circular polarization will be flipped with the flip in direction and successive passes will enhance the effect. We can use the same rotation matrix, Eq. 14.17, for the Faraday effect provided the sign is flipped after each mirror flips the circular parity.

14.2.5 Kerr Effect

An isotropic material becomes birefringent when placed in an electric field. We define n_p and n_n as the indices of light parallel and normal to the applied field. This may be viewed as a uniaxial crystal with optic axis parallel to the applied field where n_o and n_e are analogous to n_p and n_n. The difference in index, $\Delta n = n_n - n_p$ has the equation

$$\Delta n = \lambda_0 K E^2. \tag{14.19}$$

Hecht lists a number of Kerr constants for various materials. Alternatively we can represent the Kerr cell in terms of the phase retardance and the orientation of the fast axis.

The Kerr cell can be formed by sandwiching a half-wave Kerr cell between two linear polarizers at 0 and 90 degrees respectively. Incident vertically polarized light will be completely blocked when the retardance is zero and completely passed into the horizontal polarized state when proper voltage is applied to the Kerr cell. The expression for a Kerr cell blocking vertical polarization when off and converting vertical to horizontal polarization when on is represented mathematically by Eq. 14.20

$$\text{Kerr cell} = j\sin\left(\frac{\Psi}{2}\right)\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}. \tag{14.20}$$

The leading factor constant phase factor is of no consequence since it is a piston error. ψ takes on the values 0 and $\pi/2$ for off and on.

14.2.6 Pockels Effect

The Pockels effect produces retardance based on the voltage of the applied field[1],

$$\Delta\psi = 2\pi n_0^3 r_{63}\frac{V}{\lambda}, \tag{14.21}$$

where n_o is the ordinary index of refraction, r_{63} is the electro-optic constant, V is the voltage, and λ is the wavelength. The Pockels effect may be used to build a Pockels cell which works similarly to the Kerr cell. While there are important practical differences between Kerr and Pockels cells both may be modeled by Eq. 14.21.

14.3 Polarization Aberrations

In the previous section, several different types of polarization modifying components were discussed. This section addresses polarization effects in terms of polarization aberrations. This formalism is particularly suited to treating effects based on angle of incidence such as arise when beams strike refractive, reflective, or grating surfaces. In most cases these incidence angles are relatively small and the polarization affects are relatively weak. A power series development is well suited to representation of these effects. Under conditions of grazing incidence, the power series may not be a convenient representation, at least as an expansion about the normal to the surface. The polarization aberration formalism was developed by Russell Chipman several years ago and has achieved wide acceptance[2]. Chipman used a Taylor series expansion to represent effects of Fresnel reflections and thin films based on the incidence angle. In the most general form, the series expansion is

$$\mathbf{J}(i) = \sum_{k=0}^{\infty} \mathbf{J}_k i^k . \tag{14.22}$$

If the interface is isotropic, i.e., if the interface appears unchanged as it is rotated about the surface normal, only even terms need be used. Fresnel reflection is isotropic and Chipman argues that most thin film coatings are isotropic. Gratings and holograms are not isotropic with respect to the surface normal but are with respect to the grating vector, which is the important factor in determining the grating performance. A grating on a refractive or reflective surface can be treated as two separate calculations: one for diffraction and one for the surface refraction or reflection. The expansion in even powers of the incidence angle is

$$\mathbf{J}(i) = \sum_{k=0}^{\infty} \mathbf{J}_{2k} i^{2k} . \tag{14.23}$$

The incidence angle is defined to be

$$i = \cos^{-1}(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}}) . \tag{14.24}$$

Under this definition, the incidence angle is always positive. However the orientation of $\hat{\boldsymbol{p}}$ changes depending on the direction of $\hat{\boldsymbol{k}}$ if it is moved through the normal in a constant plane of incidence. The polarization properties are defined with respect to the s- and p-directions (See Chap. 11 for a discussion of the surface coordinate system). The direction of the p-vector component flips if the incident ray vector is

flipped about the normal in the plane of incidence. As mentioned above, grating properties are defined with respect to the grating unit vector, a vector normal to the grating lines. For holograms or curved gratings, the grating unit vector is defined for the local condition where the chief ray meets the grating.

14.3.1 Pauli Spin Matrices

The polarization aberrations may be developed in terms of the Pauli spin matrices

$$\mathbf{J} = \sum_{k=0}^{3} c_k \boldsymbol{\sigma}_k = \begin{bmatrix} \boldsymbol{\sigma}_0 \\ \boldsymbol{\sigma}_1 \\ \boldsymbol{\sigma}_2 \\ \boldsymbol{\sigma}_3 \end{bmatrix} \cdot \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}, \tag{14.25}$$

where the *c*'s are complex coefficients. And the Pauli spin matrices are

$$\boldsymbol{\sigma}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{\sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \boldsymbol{\sigma}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \boldsymbol{\sigma}_3 = j\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{14.26}$$

The incidence angle and direction varies across the surface. The incidence angle, considering only the geometric properties, is

$$i_x = -\frac{x}{R_x} - T_x, \; i_y = -\frac{y}{R_y} - T_y, \tag{14.27}$$

where R_x and R_y are the radii of the general toric surface and T_x and T_y are the tilt angles of the surface (measured in a counter-clockwise direction). The properties of the surface may be represented by a power series in i^2, where $i^2 = i_x^2 + i_y^2$. Consider the case of zero azimuthal angle. The zero and second orders of the power series expansion are

$$\mathbf{J} = \begin{array}{l} \left(A_{00}e^{jP_{00}} + i^2 A_{20}e^{jP_{20}}\right)\boldsymbol{\sigma}_0 \\ +\left(A_{01}e^{jP_{01}} + i^2 A_{21}e^{jP_{21}}\right)\boldsymbol{\sigma}_1 \\ +\left(A_{02}e^{jP_{02}} + i^2 A_{22}e^{jP_{22}}\right)\boldsymbol{\sigma}_2 \\ +\left(A_{03}e^{jP_{03}} + i^2 A_{23}e^{jP_{23}}\right)\boldsymbol{\sigma}_3 \end{array} = \sum_{k=0}^{3}\left(A_{0k} + i^2 A_{2k}\right)\boldsymbol{\sigma}_k. \tag{14.28}$$

If the incident angle is rotated azimuthally an angle θ, each of the Pauli spin matrices are rotated by $\mathbf{R}(\theta)\boldsymbol{\sigma}\mathbf{R}(-\theta)$ $\boldsymbol{\sigma}_0$ and $\boldsymbol{\sigma}_3$ are invariant with respect to rotation,

$$\mathbf{R}(\theta)\boldsymbol{\sigma}_1\mathbf{R}(-\theta) = \cos(2\theta)\boldsymbol{\sigma}_1 - \sin(2\theta)\boldsymbol{\sigma}_2, \tag{14.29}$$

$$\mathbf{R}(\theta)\boldsymbol{\sigma}_2\mathbf{R}(-\theta) = \sin(2\theta)\boldsymbol{\sigma}_1 + \cos(2\theta)\boldsymbol{\sigma}_2. \tag{14.30}$$

The rotated polarization matrix is

$$\mathbf{J} = \begin{array}{l} (A_{00}e^{jP_{00}} + i^2A_{20}e^{jP_{20}})\boldsymbol{\sigma}_0 \\ +(A_{01}e^{jP_{01}} + i^2A_{21}e^{jP_{21}})(\cos(2\theta)\boldsymbol{\sigma}_1 - \sin(2\theta)\boldsymbol{\sigma}_2) \\ +(A_{02}e^{jP_{02}} + i^2A_{22}e^{jP_{22}})(\sin(2\theta)\boldsymbol{\sigma}_1 + \cos(2\theta)\boldsymbol{\sigma}_2) \\ +(A_{03}e^{jP_{03}} + i^2A_{23}e^{jP_{23}})\boldsymbol{\sigma}_3 \end{array}, \tag{14.31}$$

$$\begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{array}{l} A_{00}e^{jP_{00}} + i^2A_{20}e^{jP_{20}} \\ (A_{01}e^{jP_{01}} + i^2A_{21}e^{jP_{21}})\cos(2\theta) + (A_{02}e^{jP_{02}} + i^2A_{22}e^{jP_{22}})\sin(2\theta) \\ (A_{02}e^{jP_{02}} + i^2A_{22}e^{jP_{22}})\cos(2\theta) - (A_{01}e^{jP_{01}} + i^2A_{21}e^{jP_{21}})\sin(2\theta) \\ (A_{03}e^{jP_{03}} + i^2A_{23}e^{jP_{23}}) \end{array}. \tag{14.32}$$

The trigonometric quantities are

$$\cos(2\theta) = \frac{i_y^2 - i_x^2}{i_y^2 + i_x^2}, \ \sin(2\theta) = \frac{2i_xi_y}{i_y^2 + i_x^2}. \tag{14.33}$$

Summing all $c_j\boldsymbol{\sigma}_j$

$$\mathbf{J} = \begin{bmatrix} c_0 + c_1 & c_2 - jc_3 \\ c_2 + jc_3 & c_0 - c_1 \end{bmatrix}. \tag{14.34}$$

Given the matrix J, we can compute the c-coefficients,

$$c_0 = \frac{1}{2}(j_{11} + j_{22}), \tag{14.35a}$$

$$c_1 = \frac{1}{2}(j_{11} - j_{22}), \tag{14.35b}$$

$$c_2 = \frac{1}{2}(j_{12} + j_{21}), \tag{14.35c}$$

$$c_3 = \frac{1}{2}(j_{12} - j_{21}). \tag{14.35d}$$

14.3.2 Polarization Representation in GLAD

The myriad polarization modifying effects and components can be represented by a relatively small set of polarization routines. Table 14.3 lists the current set of capabilities. The general Jones matrix allows definition of any polarization modifier. The general distributed Jones operator allows definition of any distributed polarization operation. Although any polarization component is treatable by definition of the general Jones matrix it is convenient to be able to specify many of the common devices. Linear and circular polarizers are specifiable as well as various retarders. Optical activity and Faraday rotators are specifiable. They produce a rotation of the polarization state and are differentiated by the sign of the rotation being reversed for left-handed parity in optical activity. Kerr and Pockels cells are optical switches. Polarization aberrations may be defined in terms of a power series in incident angles with general Jones matrix coefficients.

14.3.3 Polarization Effects of Surfaces

In this section the modeling of the effects of surfaces for refraction and reflection is discussed. We consider the effects due to Fresnel or single layer file coefficients at a surface. The incidence angles are computed using a combination of real and paraxial angles,

$$i_x = x\left(\frac{1}{\text{Radx}} + \frac{1}{\text{Rx}*\cos(\text{Ty})}\right) + \text{Ty}\,,\; i_y = y\left(\frac{1}{\text{Rady}} + \frac{1}{\text{Ry}*\cos(\text{Tx})}\right) + \text{Tx}\,, \tag{14.36}$$

where Radx and Rady are the radii of the reference surface of the incident beam, Rx and Ry are the radii of the surface, Tx or Ty is the tilt angle of the vertex, and i_x and i_y are the incidence angles, as defined in the description of `jsurf` in the GLAD Commands Manual. Tx and Ty may be large angles and are specified in degrees and defined to be rotation about the x- or y-axis respectively. Only one of Tx or Ty may be specified, see Fig. 14.1. The surface coordinates x and y and the surface radii Radx and Rady determine paraxial angles which are combined with the real angles Tx or Ty. In transmission a finite value of Tx or Ty results in anamorphic magnification, which is implemented by a change in units. The index is set to the substrate index at the end of propagation.

The total incidence angle is

$$i = \sqrt{i_x^2 + i_y^2}. \tag{14.37}$$

The exiting ray will be real if the incident angle is below the critical angle, or imaginary if the angle is larger than the critical angle.

Table. 14.3. Typical Jones matrices.

general Jones matrix operator	$\mathbf{J} = \begin{bmatrix} j_{11} & j_{12} \\ j_{21} & j_{22} \end{bmatrix}$
general distributed Jones operator	$e^{\mathbf{J}\Delta z}$
rotated vertical polarizer	$\mathbf{R}(\theta)\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\mathbf{R}(-\theta)$
right and left circular polarizers	$\frac{1}{2}\begin{bmatrix} 1 & j \\ -j & 1 \end{bmatrix}, \frac{1}{2}\begin{bmatrix} 1 & -j \\ j & 1 \end{bmatrix}$
rotated vertical fast axis retarder	$\mathbf{R}(\theta)\begin{bmatrix} e^{j\frac{\Psi}{2}} & 0 \\ 0 & e^{-j\frac{\Psi}{2}} \end{bmatrix}\mathbf{R}(-\theta)$
rotation due to optical activity or Faraday rotation	$\mathbf{R}(\theta_k) = \begin{bmatrix} \cos(\theta_k) & \sin(\theta_k) \\ -\sin(\theta_k) & \cos(\theta_k) \end{bmatrix}$
Kerr and Pockels cell	$j\sin\left(\frac{\Psi}{2}\right)\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$
polarization aberration	$\mathbf{J}(i) = \sum_{k=0}^{\infty} \mathbf{J}_{2k} i^{2k} \mathrm{x}$

$$\cos(i') = \begin{cases} \left(1 - \left(\frac{n}{n'}\right)^2 \sin^2(i)\right)^{1/2} & \text{for } i < i_c \text{ below critical angle} \\ j\left(\left(\frac{n}{n'}\right)^2 \sin^2(i) - 1\right)^{1/2} & \text{for } i > i_c \text{ total internal reflection} \end{cases} \tag{14.38}$$

where $i_c = \sin^{-1}((n')/n)$. The incident and exiting media are n = N1 and n' = N2, for the Fresnel surface operations. For incident angles below the critical angle, the Fresnel transmission coefficients are

$$\left(\frac{A_t}{A_i}\right)_s = t_s = \sqrt{\frac{n'}{n}}\frac{2n\cos(i)}{n\cos(i) + n'\cos(i')}, \; t_s = 0 \text{ (TIR)}, \tag{14.39a}$$

$$\left(\frac{A_t}{A_i}\right)_p = t_p = \sqrt{\frac{n'}{n}}\frac{2n\cos(i)}{n'\cos(i) + n\cos(i')}, \; t_p = 0 \text{ (TIR)}, \tag{14.40}$$

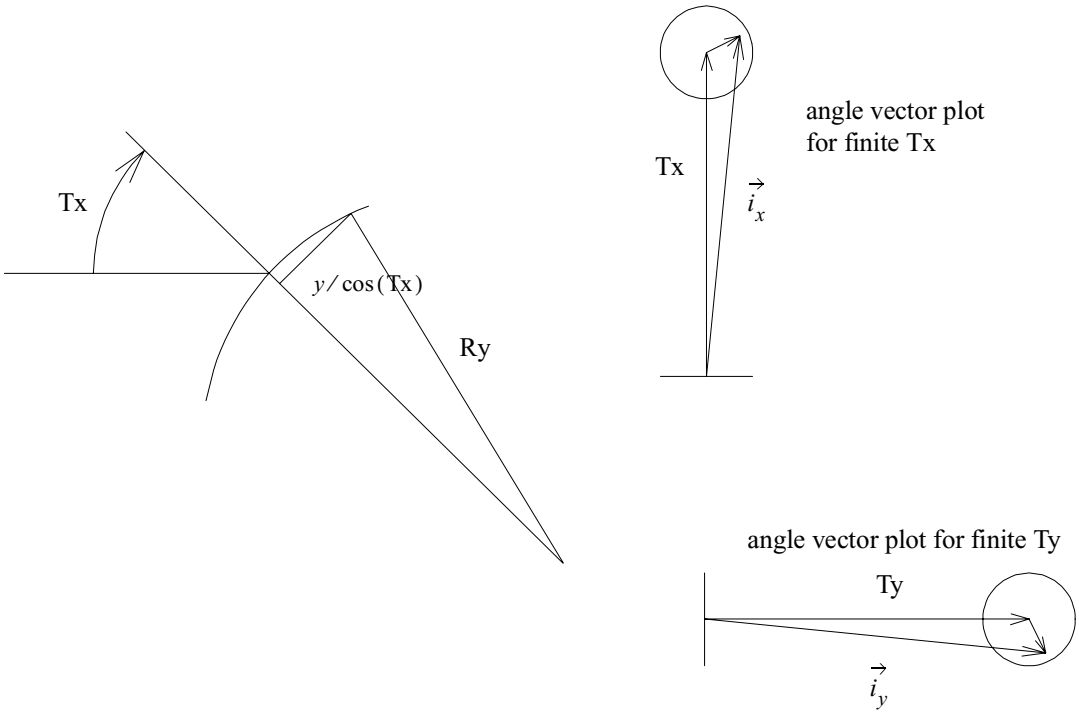


Fig. 14.1. `jsurf` allows a large tilt angle in either the x- or y-directions and paraxial angles due to surface curvature as shown by angle vector plots shown on the right for Tx (top) and Ty (bottom). The pupil is indicated by a circular region. Tx and Ty are real angles measured as rotations about the x- or y-axes. The angles due to surface curvature are treated as paraxial angles.

$$\text{x-units}' = \frac{\cos(\text{Tx}')}{\cos(\text{Tx})}\text{x-units} \quad \text{or} \quad y\text{-units}' = \frac{\cos(\text{Ty}')}{\cos(\text{Ty})}\text{y-units}, \tag{14.41}$$

The Fresnel reflection coefficients are

$$\left(\frac{A_r}{A_i}\right)_s = r_s = \frac{n\cos(i) - n'\cos(i')}{n\cos(i) + n'\cos(i')}, \tag{14.42}$$

$$\left(\frac{A_r}{A_i}\right)_p = r_p = \frac{n'\cos(i) - n\cos(i')}{n'\cos(i) + n\cos(i')}, \tag{14.43}$$

where r_s and r_p include index of refraction factors. Equations () and (14.43) are the general solutions for reflection and may be applied to ordinary reflection and total internal reflection (TIR) by including the complex value of $\cos(i')$ from Eq. (14.38). In the special case of n' being nonabsorbing, the reflection coefficients are pure phasors:

$$r_s = \exp\left\{-j2\tan^{-1}\left[\frac{n'}{n\cos(i)}\left[\left(\frac{n}{n'}\sin(i)\right)^2 - 1\right]^{1/2}\right]\right\} \text{ (TIR)}, \tag{14.44}$$

$$r_p = \exp\left\{-j2\tan^{-1}\left[\frac{n}{n'\cos(i)}\left[\left(\frac{n}{n'}\sin(i)\right)^2 - 1\right]^{1/2}\right]\right\} \text{ (TIR)}. \tag{14.45}$$

Equation () and Eq. (14.43) may also treat an absorbing medium in reflection where the irradiance absorption takes the form $\exp(-\alpha z)$. The complex index of refraction takes the form $n' \to n' + j\alpha\lambda_0/4\pi$.

Consider an example of a beam incident at Brewster's angle from air into a window with index 1.82. Brewster's angle is $i = \tan^{-1}(n'/n) = 61.2134°$. From Snell's Law the refractive angle is $i' = 28.7866°$,

$$t_s = \sqrt{\frac{n'}{n}}\frac{2n\cos(i)}{n\cos(i) + n'\cos(i')} = \sqrt{1.82}\frac{2\cos(61.2134)}{\cos(61.2134) + 1.82\cos(28.7866)} \approx 0.6257, \tag{14.46}$$

$$t_p = \sqrt{\frac{n'}{n}}\frac{2n\cos(i)}{n'\cos(i) + n\cos(i')} = \sqrt{1.82}\frac{2\cos(61.2134)}{1.82\cos(61.2134) + \cos(28.7866)} \approx 0.7412, \tag{14.47}$$

$$r_s = \frac{n\cos(i) - n'\cos(i')}{n\cos(i) + n'\cos(i')} = \frac{\cos(61.2134) - 1.82\cos(28.7866)}{\cos(61.2134) + 1.82\cos(28.7866)} \approx 0.7412, \tag{14.48}$$

$$r_p = \frac{n'\cos(i) - n\cos(i')}{n'\cos(i) + n\cos(i')} = \frac{1.82\cos(61.2134) - \cos(28.7866)}{1.82\cos(61.2134) + \cos(28.7866)} \approx 0.0. \tag{14.49}$$

For a single thin film it is necessary to solve the series solution for the multiple reflections which contribute to the steady-state reflection and transmission, as illustrated in Fig. 14.2.

$$e_{\text{refl}} = [r_{01} + t_{10}r_{12}e^{j2\beta}t_{01} + t_{10}r_{10}r_{12}^2e^{j4\beta}t_{01} + t_{10}r_{10}^2r_{12}^3e^{j6\beta}t_{01} + \ldots]e_0, \tag{14.50}$$

where

$$\beta = \frac{2\pi}{\lambda_0}n'\cos(i'), \tag{14.51}$$

and we have the relations $t_{01} = t_{10}$ and $r_{01} = -r_{10}$, allowing simplification,

$$e_{\text{refl}} = [r_{01} + t_{01}^2r_{12}e^{j2\beta} - t_{01}^2r_{10}r_{12}^2e^{j4\beta}t_{01} + t_{01}^2r_{10}^2r_{12}^3e^{j6\beta} + \ldots]e_0, \tag{14.52}$$

$$e_{\text{refl}} = [r_{01} + t_{01}^2r_{12}e^{j2\beta}[1 - r_{10}r_{12}e^{j2\beta} + r_{10}^2r_{12}^2e^{j4\beta} + \ldots]]e_0. \tag{14.53}$$

Recalling that

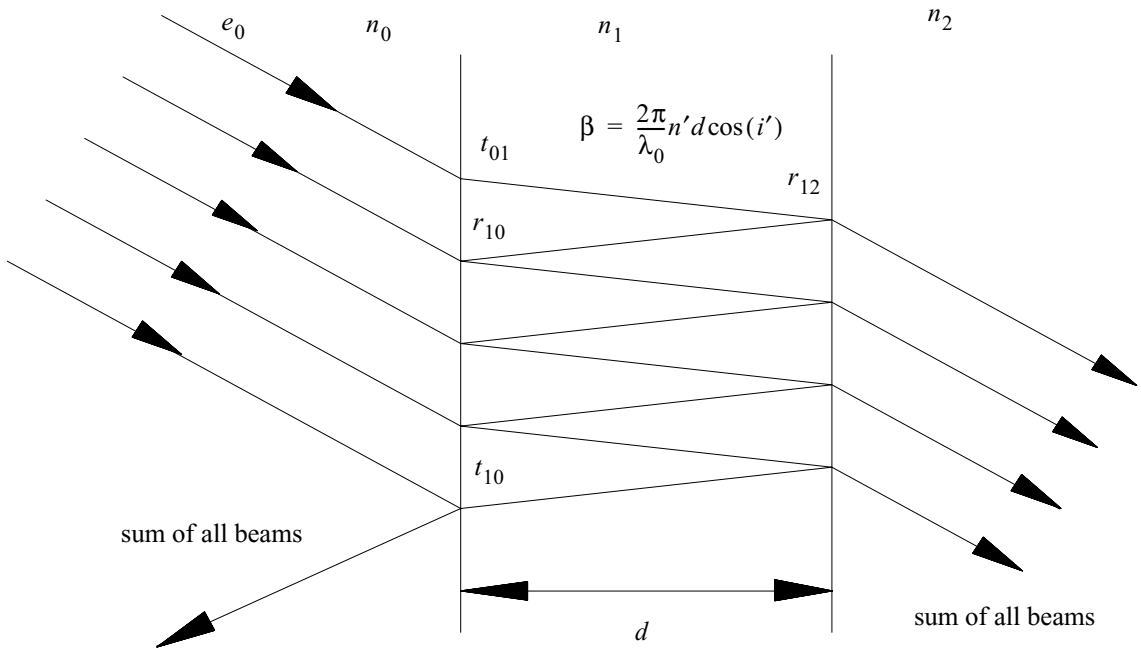


Fig. 14.2. The steady state solution for multiple reflections from a thin film separating partially reflecting surfaces. Many incident rays contribute to the amplitude of a single reflected ray.

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \ldots , \tag{14.54}$$

$$e_{\text{out}} = \frac{r_{01} + r_{12} e^{j2\beta}}{1 + r_{01} r_{12} e^{j2\beta}} e_0 \rightarrow r = \frac{r_{01} + r_{12} e^{j2\beta}}{1 + r_{01} r_{12} e^{j2\beta}} . \tag{14.55}$$

The transmission coefficients are calculated similarly,

$$e_{\text{trans}} = [t_{12} e^{j\beta} t_{01} - t_{12} r_{01} r_{12} e^{j3\beta} t_{01} + t_{12} r_{01}^2 r_{12}^2 e^{j5\beta} - \ldots] e_0 , \tag{14.56}$$

$$e_{\text{trans}} = t_{01} t_{12} e^{j\beta} [1 - r_{01} r_{12} e^{j2\beta} + r_{01}^2 r_{12}^2 e^{j4\beta} - \ldots] e_0 . \tag{14.57}$$

$$e_{\text{trans}} = \frac{t_{01} t_{12} e^{j\beta}}{1 + r_{01} r_{12} e^{j2\beta}} e_0 \rightarrow t = \frac{t_{01} t_{12} e^{j2\beta}}{1 + r_{01} r_{12} e^{j2\beta}} , \tag{14.58a}$$

$$t_s = \frac{t_{01s}t_{12s}e^{j2\beta}}{1 + r_{01s}r_{12s}e^{j2\beta}},\ t_p = \frac{t_{01p}t_{12p}e^{j2\beta}}{1 + r_{01p}r_{12p}e^{j2\beta}}, \tag{14.59}$$

$$r_s = \frac{r_{01s} + r_{12s}e^{j2\beta}}{1 + r_{01s}r_{12s}e^{j2\beta}},\ r_p = \frac{r_{01p} + r_{12p}e^{j2\beta}}{1 + r_{01p}r_{12p}e^{j2\beta}}. \tag{14.60}$$

For the single thin film surface, the equations are defined below. n_2 = Ni, n_1 = Nf, and n_0 = Ns and $\cos(i_1)$ takes real or imaginary values according to Eq. (14.38)

$$t_{01s} = \sqrt{\frac{n_1}{n_0}}\frac{2n_0\cos(i_0)}{n_0\cos(i_0) + n_1\cos(i_1)},\ t_{12s} = \sqrt{\frac{n_2}{n_1}}\frac{2n_1\cos(i_1)}{n_1\cos(i_1) + n_2\cos(i_2)}, \tag{14.61}$$

$$t_{01p} = \sqrt{\frac{n_1}{n_0}}\frac{2n_0\cos(i_0)}{n_1\cos(i_0) + n_0\cos(i_1)},\ t_{12p} = \sqrt{\frac{n_2}{n_1}}\frac{2n_1\cos(i_1)}{n_2\cos(i_1) + n_1\cos(i_2)}, \tag{14.62}$$

$$r_{01s} = \frac{n_0\cos(i_0) - n_1\cos(i_1)}{n_0\cos(i_0) + n_1\cos(i_1)},\ r_{12s} = \frac{n_1\cos(i_1) - n_2\cos(i_2)}{n_1\cos(i_1) + n_2\cos(i_2)}, \tag{14.63}$$

$$r_{01p} = \frac{n_1\cos(i_0) - n_0\cos(i_1)}{n_1\cos(i_0) + n_0\cos(i_1)},\ r_{12p} = \frac{n_2\cos(i_1) - n_1\cos(i_2)}{n_2\cos(i_1) + n_1\cos(i_2)}, \tag{14.64}$$

$$\mathbf{J}_{\text{trans}} = \begin{bmatrix} t_p & 0 \\ 0 & t_s \end{bmatrix},\ \text{x-units}' = \frac{\cos(\text{Tx}')}{\cos(\text{Tx})}\text{x-units or y-units}' = \frac{\cos(\text{Ty}')}{\cos(\text{Ty})}y\text{-units}, \tag{14.65}$$

$$\mathbf{J}_{\text{refl}} = \begin{bmatrix} r_p & 0 \\ 0 & r_s \end{bmatrix}. \tag{14.66}$$

A numerical example may be helpful. Consider a case of frustrated TIR. Light is incident at a 45º angle from an index of 2 into a thin boundary of index 1 followed by a second glass component of index 2, see Fig. 14.3. The complex value of $\cos(i')$ is selected from Eq. (14.38):

$$\text{for } n = 2,\ n' = 1,\ \text{i} = 45^\text{o},\ \text{and } d = \lambda_0/8:\ \cos(i') = j\left(\left(\frac{n}{n'}\right)^2\sin^2(i) - 1\right)^{1/2} = j, \tag{14.67}$$

$$\beta = \frac{2\pi}{\lambda_0}n'\cos(i') = j2\pi\frac{d}{\lambda_0} = j\frac{2\pi}{8} = 0.7854j,\ e^{j2\beta} = 0.2079,\ \text{and } e^{j\beta} = e^{-0.7854} = 0.4559, \tag{14.68}$$

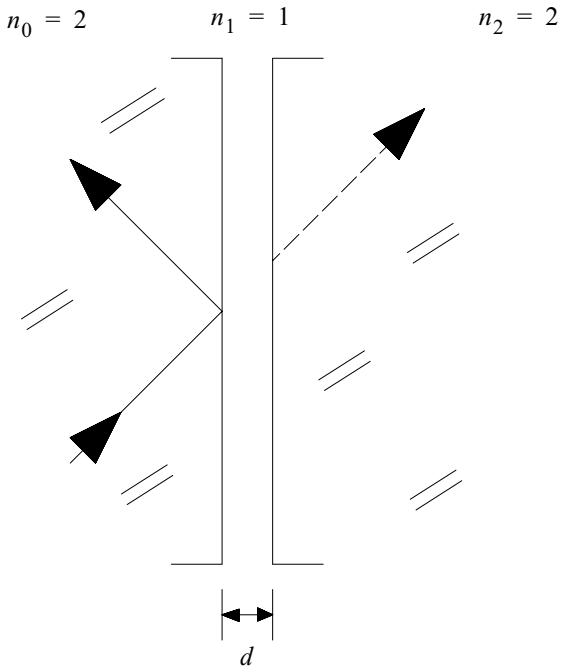


Fig. 14.3. An example of frustrated TIR. Two refractive materials of index 2 are separated by a thin boundary of index 1 and thickness d.

$$r_{01s} = \frac{n_0\cos(i_0) - n_1\cos(i_1)}{n_0\cos(i_0) + n_1\cos(i_1)} = \frac{2\cos(45) - (1)(j)}{2\cos(45) + (1)(j)} = \frac{\sqrt{2} - j}{\sqrt{2} + j} = 0.3333 - 0.9428j\,, \tag{14.69}$$

$$r_{12s} = \frac{n_1\cos(i_1) - n_2\cos(i_2)}{n_1\cos(i_1) + n_2\cos(i_2)} = \frac{j - 2\cos(45)}{j + 2\cos(45)} = \frac{j - \sqrt{2}}{j + \sqrt{2}} = -0.3333 + 0.9438j\,, \tag{14.70}$$

$$r_{01p} = \frac{n_1\cos(i_0) - n_0\cos(i_1)}{n_1\cos(i_0) + n_0\cos(i_1)} = \frac{\cos(45) - 2j}{\cos(45) + 2j} = \frac{1 - j2\sqrt{2}}{1 + 2\sqrt{2}} = -0.7778 - 0.6285j\,, \tag{14.71}$$

$$r_{12p} = \frac{n_2\cos(i_1) - n_1\cos(i_2)}{n_2\cos(i_1) + n_1\cos(i_2)} = \frac{2j - \cos(45)}{2j + \cos(45)} = \frac{j2\sqrt{2} - 1}{j2\sqrt{2} + 1} = 0.7778 + 0.6285j\,, \tag{14.72}$$

$$r_s = \frac{r_{01s} + r_{12s}e^{j2\beta}}{1 + r_{01s}r_{12s}e^{j2\beta}} = \frac{(0.3333 - 0.9428j)(1 - e^{j2\beta})}{1 - (0.3333 - 0.9428j)^2 e^{j2\beta}} = \frac{(0.3333 - 0.9428j)(0.7921)}{1 - (0.3333 - 0.9428j)^2(0.2079)}\,, \tag{14.73}$$

$$r_s = \frac{0.2640 - 0.7468j}{0.1617 + 0.1307j} = \frac{0.2091 - 0.9021j}{1.3666} = 0.1530 - 0.6601j\,, \tag{14.74}$$

$$r_p = \frac{r_{01p} + r_{12p}e^{j2\beta}}{1 + r_{01p}r_{12p}e^{j2\beta}} = \frac{(-0.7778 - 0.6285j)(1 - e^{j2\beta})}{1 - (-0.7778 - 0.6285j)^2 e^{j2\beta}} = \frac{(-0.7778 - 0.6285j)(0.7921)}{1 - (-0.2100 - 0.9777j)^2(0.2079)}, \quad (14.75)$$

$$r_p = \frac{-0.6161 - 0.4978j}{0.9563 + 0.2033j} = \frac{-0.4880 - 0.6001j}{0.9558} = -0.5106 - 0.6278j, \quad (14.76)$$

$$t_{01s} = \sqrt{\frac{n_1}{n_0}}\frac{2n_0\cos(i_0)}{n_0\cos(i_0) + n_1\cos(i_1)} = \sqrt{\frac{1}{2}}\frac{4\cos(45)}{2\cos(45) + j} = \frac{2}{\sqrt{2} + j} = \frac{2\sqrt{2} - 2j}{3}, \quad (14.77)$$

$$= 0.9428 - 0.6667j$$

$$t_{12s} = \sqrt{\frac{n_2}{n_1}}\frac{2n_1\cos(i_1)}{n_1\cos(i_1) + n_2\cos(i_2)} = \sqrt{2}\frac{2j}{j + \sqrt{2}} = \frac{2\sqrt{2} + 4j}{3} = 0.9428 + 1.333j, \quad (14.78)$$

$$t_{01p} = \sqrt{\frac{n_1}{n_0}}\frac{2n_0\cos(i_0)}{n_1\cos(i_0) + n_0\cos(i_1)} = \sqrt{\frac{1}{2}}\frac{4\cos(45)}{2\cos(45) + 2j} = \frac{2\sqrt{2}}{\sqrt{2} + j} = \frac{2\sqrt{2} - 8j}{9}, \quad (14.79)$$

$$= 0.3143 - 0.8889j$$

$$t_{12p} = \sqrt{\frac{n_2}{n_1}}\frac{2n_1\cos(i_1)}{n_2\cos(i_1) + n_1\cos(i_2)} = \sqrt{2}\frac{2j}{2j + \frac{1}{\sqrt{2}}} = \frac{8\sqrt{2} + 4j}{9} = 1.2571 + 0.444j, \quad (14.80)$$

$$t_s = \frac{t_{01s}t_{12s}e^{j2\beta}}{1 + r_{01s}r_{12s}e^{j2\beta}} = \frac{(0.9428 - 0.6667j)(0.9428 + 1.3333j)(0.4559)}{1.1617 - 0.1307j}, \quad (14.81)$$

$$t_s = \frac{0.8105 + 0.2865j}{1.1617 + 0.1307} = 0.7164 + 0.1660j, \quad (14.82)$$

$$t_p = \frac{t_{01p}t_{12p}e^{j2\beta}}{1 + r_{01p}r_{12p}e^{j2\beta}} = \frac{(0.3143 - 0.8889j)(1.2571 + 0.4444j)(0.4559)}{0.9563 - 0.2033j}, \quad (14.83)$$

$$t_p = \frac{0.3602 - 0.4458j}{0.9563 - 0.2033j} = 0.4552 - 0.3694j. \quad (14.84)$$

14.3.4 Fresnel Reflection with Absorption in the Rarer Medium

Attenuated total internal reflection (ATR) is used as a method of qualitative analysis[3]. It is necessary to generalize the refractive index from a strictly real quantity to a complex quantity where the imaginary part represents absorption $\hat{n} = n(1 + \kappa j)$. We use $+j\kappa$ so that

$$e^{j2\pi\frac{n+jn\kappa}{\lambda}z} = e^{j\frac{2\pi nz}{\lambda}} e^{-\frac{nkz}{\lambda}} \tag{14.85}$$

gives absorption. Equations () and (14.43) may be generalized to incorporate the complex refractive index

$$\text{for } \theta_i < \theta_c,\ r_s = \frac{\cos\theta_i - \sqrt{\hat{n}_{21}^2 - \sin^2\theta_i}}{\cos\theta_i + \sqrt{\hat{n}_{21}^2 - \sin^2\theta_i}}, \tag{14.86a}$$

$$\text{for } \theta_i > \theta_c,\ r_s = \frac{\cos\theta_i - j\sqrt{\sin^2\theta_i - \hat{n}_{21}^2}}{\cos\theta_i + j\sqrt{\sin^2\theta_i - \hat{n}_{21}^2}}, \tag{14.86b}$$

$$\text{for } \theta_i < \theta_c,\ r_p = \frac{\hat{n}_{21}^2\cos\theta_i - \sqrt{\hat{n}_{21}^2 - \sin^2\theta_i}}{\hat{n}_{21}^2\cos\theta_i + \sqrt{\hat{n}_{21}^2 - \sin^2\theta_i}}, \tag{14.87a}$$

$$\text{for } \theta_i > \theta_c,\ r_p = \frac{\hat{n}_{21}^2\cos\theta_i - j\sqrt{\sin^2\theta_i - \hat{n}_{21}^2}}{\hat{n}_{21}^2\cos\theta_i + j\sqrt{\sin^2\theta_i - \hat{n}_{21}^2}}, \tag{14.87b}$$

where $\hat{n}_{21} = \hat{n}_2/n_1$ and we have considered only strictly real values of the incident refractive index n_1.

For $n_1 = 2$, $n_2 = 1$, $\theta_i = 0$, $\kappa = 0$

$$r_s = -r_p = \frac{n_1 - n_2}{n_1 + n_2} = \frac{1}{3}. \tag{14.88}$$

For $n_1 = 2$, $n_2 = 1$, $\theta_i = 45°$ (y-tilt), $\kappa = 0$

$$r_s = \frac{\cos\theta_i - j\sqrt{\sin^2\theta_i - \hat{n}_{21}^2}}{\cos\theta_i + j\sqrt{\sin^2\theta_i - \hat{n}_{21}^2}} = \frac{1}{3} - \frac{2\sqrt{2}}{3}j \text{ and } r_s = \frac{\hat{n}_{21}^2\cos\theta_i - j\sqrt{\sin^2\theta_i - \hat{n}_{21}^2}}{\hat{n}_{21}^2\cos\theta_i + j\sqrt{\sin^2\theta_i - \hat{n}_{21}^2}} = \frac{7}{9} - \frac{4\sqrt{2}}{9}j. \tag{14.89}$$

For $n_1 = 2$, $n_2 = 1 + j$, $\theta_i = 0$ (y-tilt), $\kappa = 1 + j$

$$r_s = -r_p = \frac{n_1 - n_2 - n_2\kappa j}{n_1 + n_2 + n_2\kappa j} = 0.2 - 0.4j\,. \tag{14.90}$$

For $n_1 = 2$, $n_2 = 1 + j$, $\theta_i = 45°$ (y-tilt), $\kappa = 1$

$$r_s = \frac{\cos\theta_i - j\sqrt{\sin^2\theta_i - \hat{n}_{21}^2}}{\cos\theta_i + j\sqrt{\sin^2\theta_i - \hat{n}_{21}^2}} = \frac{1 - j\sqrt{1-j}}{1 + j\sqrt{1-j}} = \frac{1 - 2^{1/4}\left(\sin\frac{\pi}{8} + j\cos\frac{\pi}{8}\right)}{1 + 2^{1/4}\left(\sin\frac{\pi}{8} + j\cos\frac{\pi}{8}\right)} = -0.125 - 0.661j\,, \tag{14.91}$$

$$r_p = \frac{\hat{n}_{21}^2\cos\theta_i - j\sqrt{\sin^2\theta_i - \hat{n}_{21}^2}}{\hat{n}_{21}^2\cos\theta_i + j\sqrt{\sin^2\theta_i - \hat{n}_{21}^2}} = \frac{\frac{1}{2} - \sqrt{1-j}}{\frac{1}{2} + \sqrt{1-j}} = \frac{1 - 2^{1/4}\left(\cos\frac{\pi}{8} - \sin\frac{\pi}{8}\right)}{1 + 2^{1/4}\left(\cos\frac{\pi}{8} - \sin\frac{\pi}{8}\right)} = -0.421 + 0.165j\,. \tag{14.92}$$

We shall consider solution of Eqs (14.86a–14.86b) and (14.87a–14.87b) for the two cases: $\theta_i < \theta_c$ (– solutions) and $\theta_i > \theta_c$ (+ solutions), where $\theta_c = \sin^{-1}(n_1/n_2)$. We introduce the variables ξ, η, u, and v for convenience.

$$r_s = \frac{\xi - \eta_\pm}{\xi + \eta_\pm}\,, \tag{14.93a}$$

where

$$\xi = \cos^2\theta_i + (u^2 + v^2)^{1/2}\,, \tag{14.93b}$$

$$\eta_\pm = \pm\sqrt{2}\cos(\theta_i)[(u^2 + v^2)^{1/2} \mp \mathrm{v}]^{1/2}\,, \tag{14.93c}$$

$$u = \sin^2(\theta_i) - n_{21}^2 + n_{21}^2\kappa^2\,, \tag{14.93d}$$

$$v = 2n_{21}^2\kappa^2\,, \tag{14.93e}$$

$$r_p = \frac{\Xi - H_\pm}{\Xi + H_\pm}\,, \tag{14.94a}$$

where

$$\Xi = \cos^2(\theta_i)\left\{[\sin^2(\theta_i) - u^2]^2 + v^2\right\} + (u^2 + v^2)^{1/2}, \tag{14.94b}$$

$$H_{\pm} = \pm 2\cos(\theta_i)\left\{\frac{n_{21}^2}{\sqrt{2}}(1-\kappa^2)[(u^2+v^2)^{1/2} \mp u]^{1/2} \pm \frac{v}{\sqrt{2}}[(u^2+v^2)^{1/2} \pm u]^{1/2}\right\}. \tag{14.94c}$$

14.3.5 Goos-Hanchen effect

In the case of small diameter beams incident on a TIR interface, the beam experiences a translation along the surface. Figure 14.4 illustrates schematically the translation. This may be understood as simply due to the phase shift variation with angle associated with reflection at TIR from Eq. () and (14.43). The curved phase of TIR is illustrated in Fig. 14.5 as a curve response versus angle. The linear term of the phase function produces the Goos Hanchen shift. The second order term (curvature) produces a compression or expansion of the beam depending on the sign of the term.

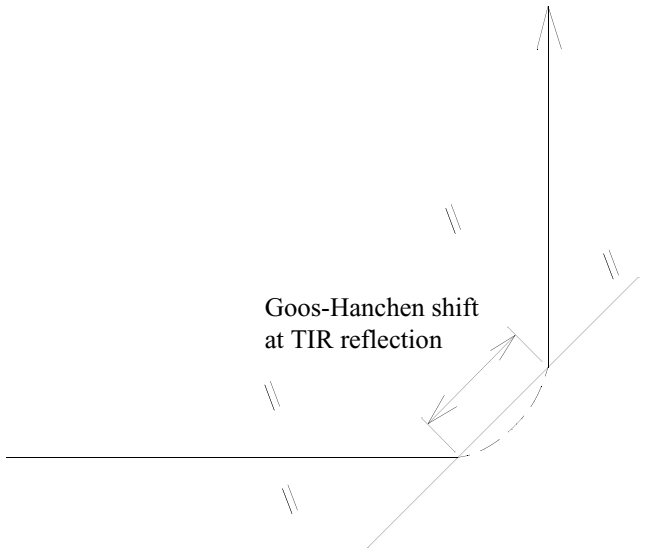


Fig. 14.4. Goos-Hanchen shift. A small diameter beam incident from the left is incident on a boundary with a lower index medium, making a TIR reflection. The beam is slightly shifted and also experiences some distortion due to the variation of phase of the reflection coefficients with angle. This is called the Goos-Hanchen shift.

References

1. Eugene Hecht, *Optics*, 2nd Ed., Addison-Wesley (1987).
2. R. Chipman, “Polarization Aberrations”, PhD Dissertation, University of Arizona (1977).
3. G. Muller, K. Abraham, and M. Schaldach, “Quantitative ATR spectroscopy: some basic considerations”. Appl. Opt. Vol. 20, No. 7, pp1182-1190 (1981).

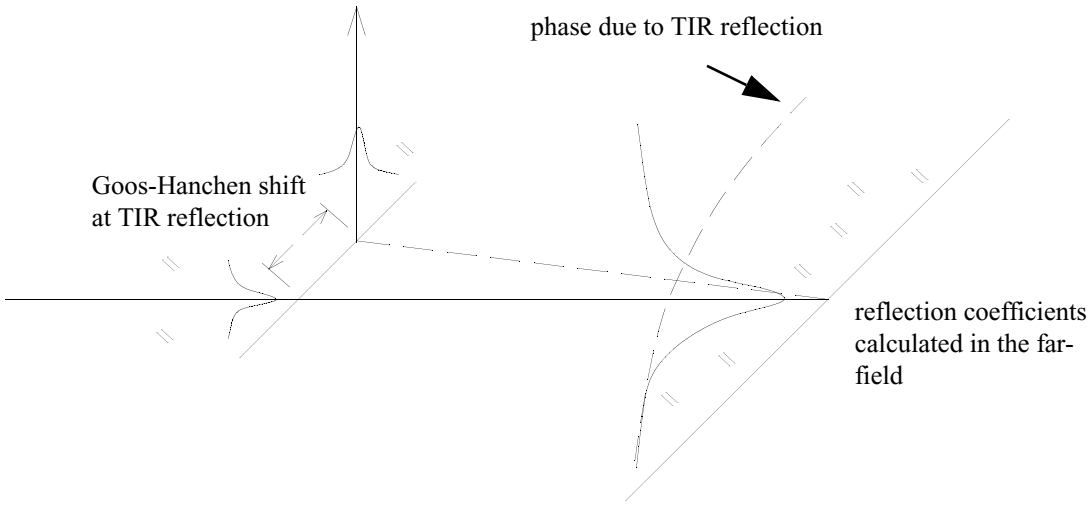


Fig. 14.5. Goos-Hanchen shift, far-field implementation. A small diameter beam incident from the left is incident on a boundary with a lower index medium, making a TIR reflection. For a flat surface, the Fresnel reflection coefficients may be implemented in the far-field. The figure on the right shows the beam expanded in the far-field. The `jsurf` command implements the Fresnel reflection coefficients for the local angles. The variation of phase with angle creates a net linear phase variation across the far-field beam as well as distortion terms. The back-propagated beam is shifted as well as distorted, giving the Goos-Hanchen shift.

15. Waveguide Grating Couplers

In the previous chapters, radiation has been assumed to propagate in isotropic media or, in the case of certain polarization properties, nearly isotropic media. In free-space or homogeneous media, the eigenmodes are plane waves. Fourier optics, based on plane wave decomposition, are powerful methods for numerical calculations of propagation. Waveguides, have different modes based on the indices of substrate, film, and cover materials, and the thickness of the waveguides. Each waveguide mode can be propagated by applying a complex eigenvalue for propagation,

$$\psi_n(z) = \psi_n(0) e^{j\beta_n z}. \tag{15.1}$$

The eigenvalue $e^{j\beta_n z}$ is similar to the eigenvalues in free-space e^{jkz}, with β representing the equivalent wavenumber in the waveguide. If the waveguide supports several waveguide modes with different propagation constants, the complex amplitude will vary in the course of optical propagation. The axial variation of complex amplitude takes the form

$$\phi(z) = \sum_{n=1}^{N} c_n e^{j\beta_n z} \psi_n(0), \tag{15.2}$$

where the coefficients c_n specify the mode composition. In general, the β_n are not identical, so $\phi(z)$ varies with z. For high quality optical waveguides it is desirable to build the device so that only one waveguide mode is supported, giving a smoothly varying phase function. The current treatment will be limited to single-mode waveguides.

Slab waveguides represent an important class of devices. These devices consist of a planar substrate with a waveguide layer on the top of the substrate as illustrated in Fig. 15.1.

In the direction perpendicular to the surface normal, the radiation will propagate as a guided mode. In directions parallel to the surface, the radiation will propagate according to free-space theory. If only one waveguide mode is supported, the problem is particularly easy because the guided mode is invariant except for the phase evolution due to the optical path. The propagation constant of the waveguide β is determined by the indices of the cover layer n_s, film n_f, and substrate n_s and the thickness. An effective index for the guided propagation n_0 may be defined such that $\beta = 2\pi n_0/\lambda$. The index to be used for free-space diffraction in the in-plane direction is not necessarily the same as n_0, but in most applications the difference may be neglected. Waveguide grating couplers allow guided modes to be outcoupled to radiated modes and radiated modes to be incoupled to guided modes. A particularly important class of grating couplers are the focusing couplers. These devices outcoupled to (or incoupled from) a point focus. The frequency and direction of the grating will vary across the grating area so that the light is diffracted at each point of the grating toward the focus point.

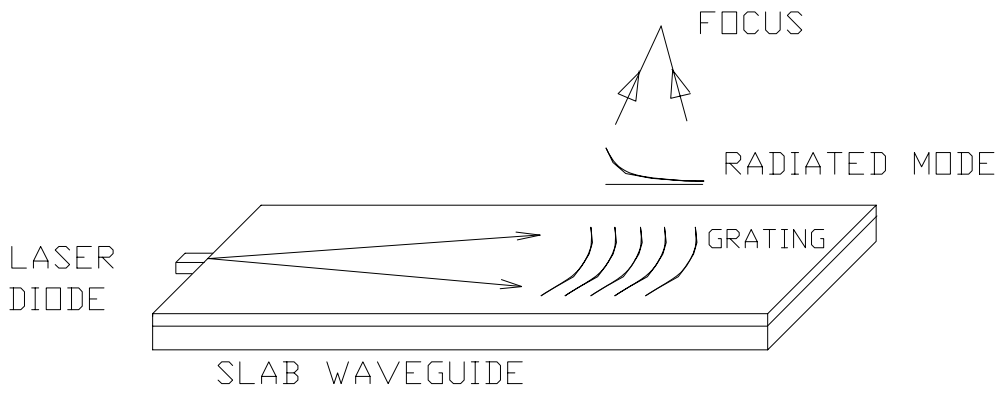


Fig. 15.1. Schematic of grating coupler as used in optical data storage. Light from a laser diode is injected into a slab waveguide. The light propagates as a guided mode in the vertical direction — trapped in a boundary layer of higher index. A grating of variable index or surface relief is put on the waveguide. The grating lines are curved to form a diffracted beam which converges to the focus. The guided mode decays as it passes under the grating, as the light is scattered out.

15.1 Perturbation Model of Grating Coupling

The focus of our attention in this chapter is the incoupling and outcoupling of radiation due to gratings on the surface of the waveguide. The configuration and coordinate system is illustrated in Fig. 15.2. Considerable simplification of the theory is possible if the grating is considered to be weak. The grating may be represented as a perturbation of the electric permitivity of the form

$$\varepsilon_p e^{j\mathbf{\Gamma}\cdot\boldsymbol{r}}. \tag{15.3}$$

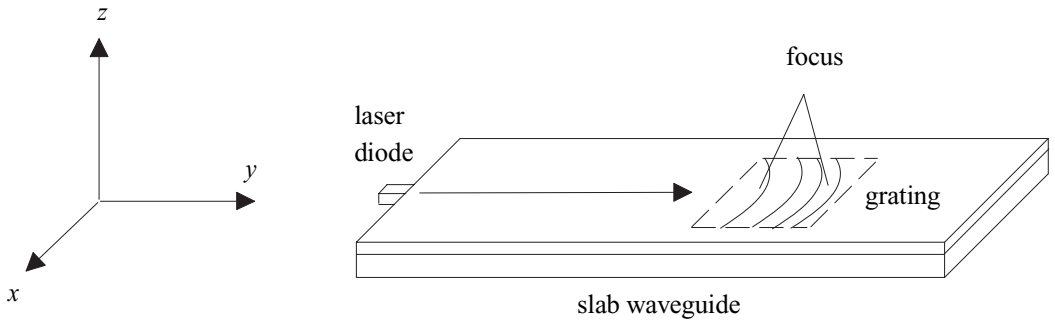


Fig. 15.2. Slab waveguide consisting of a planar substrate surface with a thin, high index layer on the top. The light will propagate as a guided wave when considered in a direction parallel to the surface normal and according to free-space propagation in the direction parallel to the surface.

We shall first solve the problem for ideal linear gratings. Focusing gratings, which have variable grating period and direction, can be considered to be a combination of linear gratings, so the nonnormal, constant period case will apply to focusing devices. For typical grating fabrication methods and weak gratings, ε_p will be a relatively thin layer on the top of the waveguide. Following the approach taken by Shiau, we write the E&M fields in the perturbed waveguide

$$\boldsymbol{E} = \boldsymbol{E}_u + \boldsymbol{E}_p, \tag{15.4}$$

$$\boldsymbol{H} = \boldsymbol{H}_u + \boldsymbol{H}_p, \tag{15.5}$$

where the subscripts *u* and *p* indicate the unperturbed and perturbed fields. Writing the fields (ignoring the time phase factor which is identical for all terms), we have

$$\nabla\times(\boldsymbol{E}_u + \boldsymbol{E}_p) = j\omega\mu_0(\boldsymbol{H}_u + \boldsymbol{H}_p), \tag{15.6}$$

$$\nabla\times(\boldsymbol{H}_u + \boldsymbol{H}_p) = -j\omega\varepsilon_0(\varepsilon_u + \varepsilon_p)(\boldsymbol{E}_u + \boldsymbol{E}_p). \tag{15.7}$$

The unperturbed fields satisfy the equations,

$$\nabla\times\boldsymbol{E}_u = j\omega\mu_0\boldsymbol{H}_u, \tag{15.8}$$

$$\nabla\times\boldsymbol{H}_u = -j\omega\varepsilon_0\varepsilon_u\boldsymbol{E}_u. \tag{15.9}$$

By making use of the unperturbed solutions, we have the simplifications

$$\nabla\times\boldsymbol{E}_p = j\omega\mu_0\boldsymbol{H}_p, \tag{15.10}$$

$$\nabla\times\boldsymbol{H}_p = -j\omega\varepsilon_0(\varepsilon_u\boldsymbol{E}_p + \varepsilon_p\boldsymbol{E}_p + \boldsymbol{P}), \tag{15.11}$$

where

$$\boldsymbol{P} = \varepsilon_p\boldsymbol{E}_u. \tag{15.12}$$

The term $\varepsilon_p\boldsymbol{E}_p$ may be neglected for weak gratings. Combining Eqs. (15.10) and (15.11) we have the equation,

$$\nabla\times\nabla\times\boldsymbol{E}_p = \omega\varepsilon_0\mu_0(\varepsilon_u\boldsymbol{E}_p + \boldsymbol{P}). \tag{15.13}$$

The right hand side of Eq. (15.13) can be seen to consist of a diffraction term $\boldsymbol{E}_p$ and a source term $\boldsymbol{P}$. In the absence of the source term, Eq. (15.13) represents diffraction of the field due to the perturbation. The source term may be included by means of a variety of theoretical approaches.

$$\nabla\times\nabla\times\boldsymbol{E}_p \;=\; \nabla^2\boldsymbol{E}_p - \nabla(\nabla\cdot\boldsymbol{E}_p) \;=\; \omega\varepsilon_0\mu_0(\varepsilon_u\boldsymbol{E}_p + \boldsymbol{P})\,, \tag{15.14}$$

$$\nabla\cdot\boldsymbol{E}_p \;=\; -\frac{1}{\varepsilon_p}\nabla\varepsilon_p\cdot\boldsymbol{E}_u\,, \tag{15.15}$$

$$\nabla^2\boldsymbol{E}_p \;=\; -k_0^2\varepsilon_u\boldsymbol{E}_p - \frac{1}{\varepsilon_u}\nabla(\nabla\varepsilon_u\cdot\boldsymbol{E}_u) - k_0^2\varepsilon_p\boldsymbol{E}_u\,, \tag{15.16}$$

$$\varepsilon_p(x,y,z) \;=\; \varepsilon_p\mathrm{rect}\left(\frac{z}{d}\right)e^{j\boldsymbol{\Gamma}\cdot\boldsymbol{r}} \;=\; \varepsilon_p\mathrm{rect}\left(\frac{z}{d}\right)e^{j(\Gamma_x x + \Gamma_y y)}\,, \tag{15.17}$$

where $\mathrm{rect}(z/d)$ indicates that the perturbation to permitivity is limited to the grating region of width d.

$$\nabla(\varepsilon_p e^{j(\Gamma_x x + \Gamma_y y)}) \;=\; \varepsilon_p(j\Gamma_x\hat{\boldsymbol{x}} + j\Gamma_x\hat{\boldsymbol{y}})e^{j(\Gamma_x x + \Gamma_y y)}\,. \tag{15.18}$$

Taking the m^{th} order.

$$(\nabla\varepsilon_p\cdot\boldsymbol{E}_u) \;=\; \varepsilon_p(jm\Gamma_x\hat{\boldsymbol{x}} + jm\Gamma_x\hat{\boldsymbol{y}})e^{j(m\Gamma_x x + m\Gamma_y y)}\cdot\boldsymbol{E}_u e^{j\beta y}\hat{\boldsymbol{x}}\,, \tag{15.19}$$

$$\nabla(\nabla\varepsilon_p\cdot\boldsymbol{E}_u) \;=\; \varepsilon_p jm\Gamma_x(jm\Gamma_x\hat{\boldsymbol{x}} + j(m\Gamma_x + \beta)\hat{\boldsymbol{y}})e^{j(m\Gamma_x x + (m\Gamma_y + \beta)y)}\boldsymbol{E}_u\,, \tag{15.20}$$

$$\nabla(\nabla\varepsilon_p\cdot\boldsymbol{E}_u) \;=\; -\varepsilon_p(m^2\Gamma_x^2\hat{\boldsymbol{x}} + j(m^2\Gamma_x\Gamma_y + \beta m\Gamma_x)\hat{\boldsymbol{y}})e^{j(m\Gamma_x x + (m\Gamma_y + \beta)y)}\boldsymbol{E}_u\,, \tag{15.21}$$

$$\nabla(\nabla\varepsilon_p\cdot\boldsymbol{E}_u) \;=\; -\varepsilon_p m^2\Gamma^2\boldsymbol{E}_u\left[\sin^2\theta\hat{\boldsymbol{x}} + \left(\sin\theta\cos\theta + \sin\theta\frac{\beta}{m\Gamma}\right)\hat{\boldsymbol{y}}\right]e^{j(m\Gamma_x x + (m\Gamma_y + \beta)y)}\,. \tag{15.22}$$

After removing the spatial phase factors by considering only the Bragg condition.

$$\nabla^2\boldsymbol{E}_p \;=\; -k_0^2\varepsilon_u\boldsymbol{E}_p \pm \frac{1}{\varepsilon}\varepsilon_p m^2\Gamma^2 E_u\left[\sin^2\theta\hat{\boldsymbol{x}} + \left(\sin\theta\cos\theta + \sin\theta\frac{\beta}{m\Gamma}\right)\hat{\boldsymbol{y}}\right] - k_0^2\varepsilon_u\boldsymbol{E}_p\,, \tag{15.23}$$

$$\nabla^2\boldsymbol{E}_p \;=\; -k_0^2\varepsilon_u\boldsymbol{E}_p + \frac{\varepsilon_p}{\varepsilon_u}\boldsymbol{E}_u[(m^2\Gamma^2\sin^2\theta - \varepsilon_u k_0^2)\hat{\boldsymbol{x}} + (m^2\Gamma^2\sin\theta\cos\theta + m\Gamma\beta\sin\theta)\hat{\boldsymbol{y}}]\,. \tag{15.24}$$

The source term $\boldsymbol{P}$ for the perturbation model, is driven entirely by the unperturbed electric field with scalar modulation determined by ε_p. We shall consider both an induced dipole model and coupled mode model to determine the characteristics of the perturbation field.

15.2 Induced Dipole

The source term of Eq. (15.13) can be considered to be a coherent object consisting of a field of oscillating dipoles. The spatial phase at each point is determined by the propagation constants and direction for guided mode, grating, and radiated mode. For the linear grating the spatial phase factors are $\exp(j\boldsymbol{\beta}\cdot\boldsymbol{r})$, $\exp(j\boldsymbol{\Gamma}\cdot\boldsymbol{r})$, and $\exp(j\boldsymbol{k}\cdot\boldsymbol{r})$ for the guided mode, grating, and radiated mode respectively. Both the guided mode and the grating are parallel to the plane of the surface, i.e.,

$$\boldsymbol{\beta}\cdot\hat{\boldsymbol{n}} = 0, \tag{15.25}$$

$$\boldsymbol{\Gamma}\cdot\hat{\boldsymbol{n}} = 0. \tag{15.26}$$

Grouping the spatial phase factors in Equation (15.13), and requiring phase matching on the surface of the waveguide, we have the Bragg condition

$$\hat{\boldsymbol{n}}\times(\boldsymbol{\beta}-\boldsymbol{k}) = m(\hat{\boldsymbol{n}}\times\boldsymbol{\Gamma}). \tag{15.27}$$

By considering stationary phase, we can see that the only direction having significant radiation will be the one determined by the Bragg condition. We shall assume the Bragg condition is met for outcoupling and apply a mode-matching condition for input to take into account detuning. When viewing the light scattered out of the waveguide from a direction satisfying the Bragg condition, we would see a coherent source of constant phase. The radiation of light from a coherent source is well-understood. One approach is to consider the radiation at a distant point to be due to dipole radiation from each point on the surface. If our observation point is a large distance from the surface, the radiation due to each point is (see Jackson[1])

$$\boldsymbol{E}\propto\frac{\hat{\boldsymbol{k}}\times\boldsymbol{p}\times\hat{\boldsymbol{k}}}{r}, \tag{15.28}$$

where $\boldsymbol{p}$ is the dipole moment and r is the distance from the observation point to the dipole. According to Eq. (15.27) $\boldsymbol{E}$ is driven by the projection of the dipole onto the transverse plane of the observation direction and the polarization of $\boldsymbol{E}$ is determined by the projected character of the dipole. For a linear grating, a collimated beam is generated and one must be at infinite distance to observe the surface of the grating while satisfying the Bragg condition. For a focusing grating, an observation point at the focus position will satisfy the Bragg condition when viewing all points on the surface. We may, therefore, calculate the complex amplitude at the center of focus of a focusing grating, by calculating Eq. (15.27) for all points on the surface of the waveguide. To be more precise we must consider $\boldsymbol{P}$ as a volume source with the further complication that the dipoles lie in a medium of different index.

Figure 15.3a illustrates the fact that an ideal dipole in vacuum generates radiation (in the far-field, i.e., $r\gg\lambda$) based on the projection of the dipole oscillation onto a plane transverse to the direction of observation. Fig. 15.3b illustrates dipole radiation as observed after crossing through a refractive boundary. The situation is little changed from the vacuum situation. There are Fresnel transmission coefficients which must be applied to the radiation at the boundary. Similarly, when propagating from a high index medium to a lower one we have refraction effects and Fresnel transmission coefficients to consider, as shown in Fig.

15.3c. The Fresnel transmission coefficients will have a small effect on the observed polarization by transmitting s- and p-polarization differently. For near-normal incidence, the effect on polarization is very small and we may safely say that the polarization of the radiation is simply the projection of the dipole on the transverse plane.

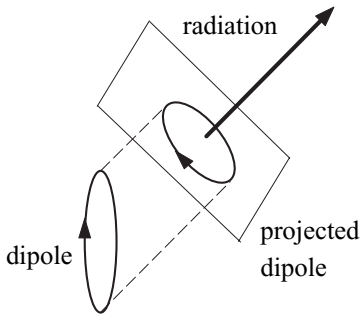


Fig. 15.3a. Dipole in vacuum projects onto transverse plane.

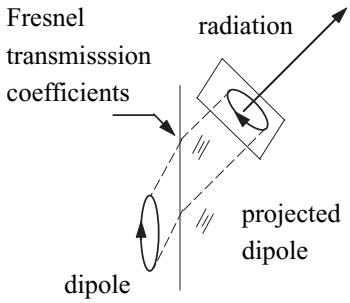


Fig. 15.3b. Dipole in vacuum projects onto transverse plane in vacuum after refraction and Fresnel losses.

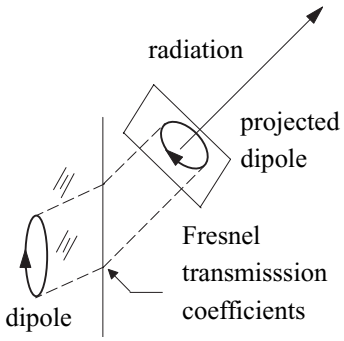


Fig. 15.3c. Dipole in high index material projects onto transverse plane in vacuum after refraction and Fresnel losses

For the focusing grating and taking the observation point at the center of the focused beam, we can determine the complex amplitude field and polarization of the induced dipoles. The induced dipole field must meet the Bragg condition and the spatial restrictions of $\varepsilon_p(z)\boldsymbol{E}_u(z)$. The perturbation to the permitivity is generally localized to a thin layer on the top of the grating. We must perform a volume integration over the waveguide region

$$\boldsymbol{E}_p \propto \frac{1}{r}\int \varepsilon(x,y,z)(\hat{\boldsymbol{k}}\times\boldsymbol{p}\times\hat{\boldsymbol{k}})dxdydx\,, \tag{15.29}$$

$$\boldsymbol{E}_p \propto \frac{1}{r}\alpha(x,y)\tilde{E}_u(x,y)(\hat{\boldsymbol{k}}\times\boldsymbol{p}\times\hat{\boldsymbol{k}})\,, \tag{15.30}$$

where

$$\alpha(x,y) = \frac{\int \varepsilon_p(x,y,z)\boldsymbol{E}_u(x,y,z)dz}{\sqrt{\int \boldsymbol{E}_u^{\dagger}\cdot\boldsymbol{E}_u dz}}\,, \tag{15.31}$$

and $\tilde{\boldsymbol{E}}$ is defined to be the integrated complex amplitude taken along the z-direction such that—†

$$\left|\tilde{\boldsymbol{E}}_u\right|^2 = \int \boldsymbol{E}_u^{\dagger}\cdot\boldsymbol{E}_u dz\ . \tag{15.32}$$

For a grating of constant modulation depth, $\varepsilon_p(x, y, z)$ is a constant and the integration to form α is also independent of the (x, y) coordinates of the pupil. In that case, the only variation is due to the projection of the dipole unit vector $\hat{\boldsymbol{p}}$ onto the direction of observation. Consequently, the complex amplitude and polarization properties in the pupil of a focusing grating are readily determined from the unperturbed properties of the guided mode field and the geometry of the focusing beam.

The properties of the guided electric field have the forms (Yariv, Quantum Electronics, [Ref. 2, p512-514]),

TE mode

$$ {}_x(z) = \begin{cases} C\exp(-qz) & 0 < z < \infty \\ C\left[\cos(hz) - \dfrac{q}{h}\sin(hz)\right] & -t < z < 0 \\ C\left[\cos(ht) + \dfrac{q}{h}\sin(ht)\right]\exp[p(z+t)] & -\infty < z < -t\,, \end{cases} \tag{15.33} $$

$$ \begin{aligned} h &= \left(n_f^2 - \beta^2\right)^{1/2} \\ q &= \left(\beta^2 - n_c^2 k^2\right)^{1/2} \\ p &= \left(\beta^2 - n_s^2 k^2\right)^{1/2}. \end{aligned} \tag{15.34} $$

TM mode

$$ e_x(z) = \begin{cases} -C\left[\dfrac{h}{q}\cos(ht) + \sin(ht)\right]\exp[p(z+t)] & z < -t \\ C\left[-\dfrac{h}{q}\cos(hz) + \sin(hz)\right] & -t < z < 0 \\ -C\dfrac{h}{q}\exp(-qz) & z > 0\,, \end{cases} \tag{15.35} $$

$$ e_z(z) = \frac{\beta}{\omega\varepsilon} h_x(z)\,, \tag{15.36} $$

$$ e_y(z) = -\frac{j}{\omega\varepsilon}\frac{\partial}{\partial x} h_x(z)\,. \tag{15.37} $$

Equation (15.33) gives the variation as a function of depth and Eqs. (15.36) and (15.37) give the depth variation for TM and a sketch of the appearance of the TE and TM guided mode fields.

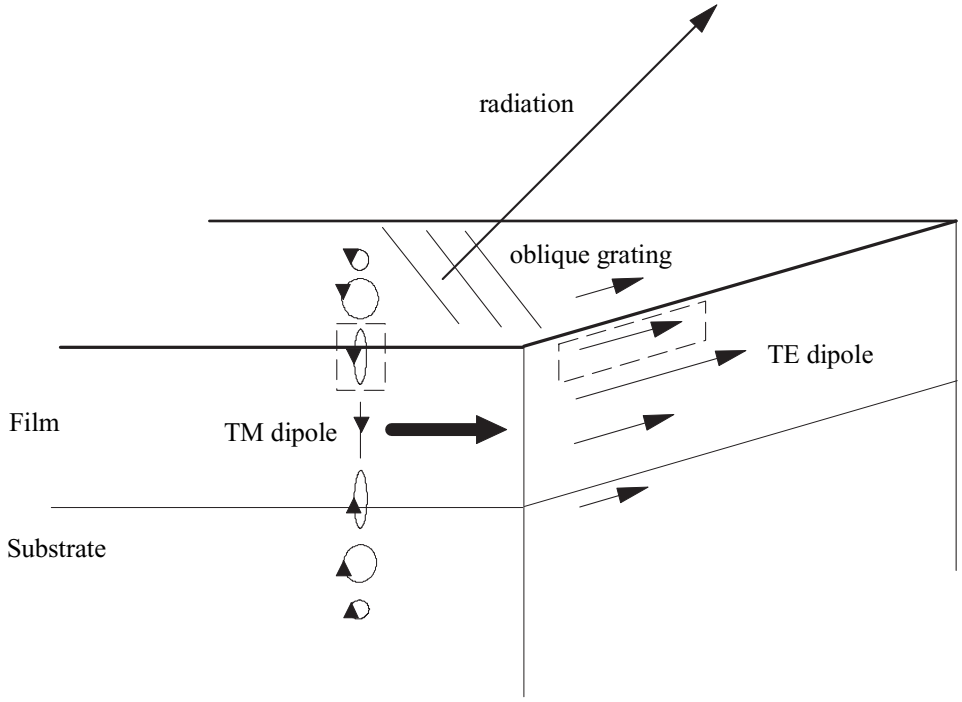


Fig. 15.4. Schematic of TE and TM dipoles in waveguide outcoupling. TE has a well-defined linear polarization at all levels and generates linearly polarized output. TM is elliptically polarized with variation as a function of level. The elliptical polarization of TM at the grating level projects onto the transverse plane of the radiation giving different states of elliptical polarization across the aperture of a focusing grating coupler. The grating is considered to have insignificant surface relief so the surface normal is constant across the surface.

The TE mode is simply a linearly polarized field, as illustrated in Fig. 15.4 by a series of vectors parallel to the surface, transverse to the direction of propagation and varying in amplitude according to depth. The dipole projection of this uniform linearly polarized field must create a field in the pupil of a focusing grating which is also linearly polarized. Figure 15.5.a illustrates the exponential decay of energy which is typical of focusing gratings. The TE mode is propagating in the –y direction.

Fig. 15.6.a shows the elliptical polarization map for TE indicating perfectly uniform polarization. We are neglecting the slight effects of Fresnel transmission coefficients which depend on the local s- and p-polarization components as the field exits from material into air. TM polarization exhibits components parallel to the surface normal and in the direction of propagation. These components differ by 90º of phase. The TM polarization may be considered to be elliptical polarization when viewed perpendicular to the direction of propagation, as shown in Fig. 15.4. Only the elliptical polarization at the grating level contribute to the radiation. By the induced dipole argument, the polarization in the pupil is simply the projection of the dipoles in the grating layer onto the transverse plane associated with the direction of propagation.

The most important factor for the outcoupling of TM guided radiation is the geometric projection of the dipole. The projected dipoles determine both the orientation of polarization and the coupling coefficient. Fig. 15.5b illustrates the intensity profile of the outcoupled radiation. The dip in the center is due to the fact that the vertical component of the TM polarization does not contribute to radiation in the vertical direction.

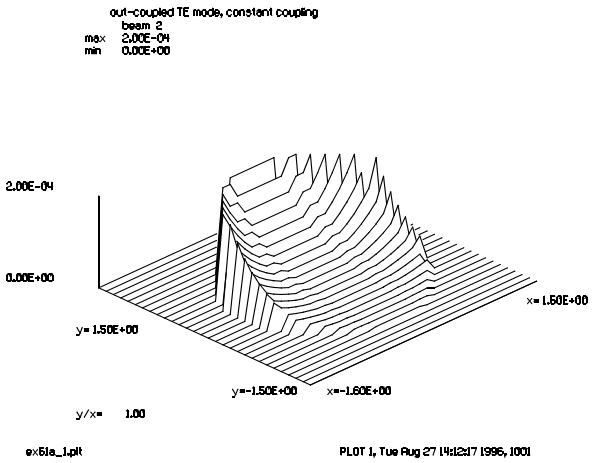


Fig. 15.5.a. Isometric plot of TE radiation output. The guided mode is propagating down from the top. The pupil exhibits simple exponential decay. The grating is circular, causing the outer edges to begin the exponential decay later than the center.

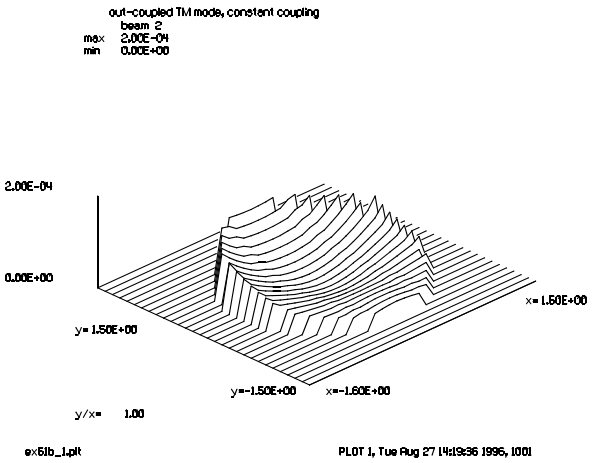


Fig. 15.5b. Isometric plot of TM radiation showing pronounced departure from the exponential decay. The outcoupling is much lower in the center because the vertical component of TM does not contribute.

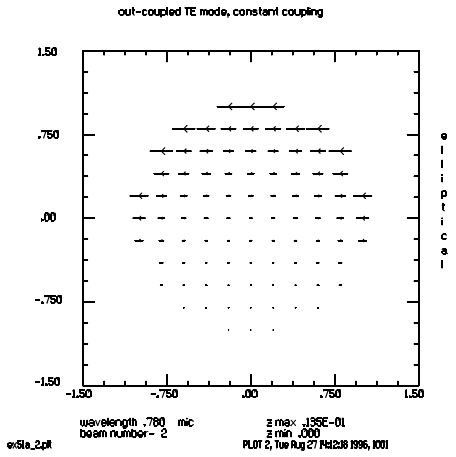


Fig. 15.6.a. Elliptical polarization plot of TM radiation showing pronounced departure from the exponential decay. The outcoupling is much lower in the center because the vertical component of TM does not contribute.

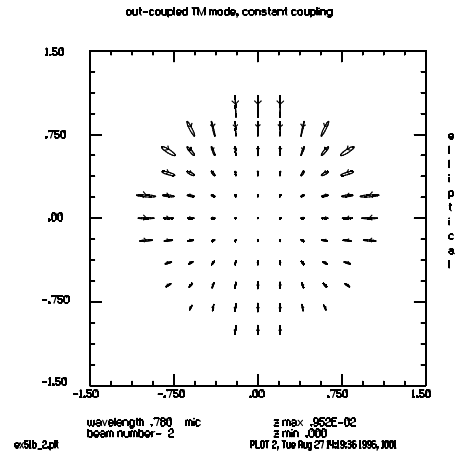


Fig. 15.6b. The polarization varies across the pupil of a focusing grating for TM because the elliptical dipoles are viewed from different directions as seen from the focus point.

Fig. 15.6b illustrates the polarization in the pupil due to radiation scattered from the TM mode. Figures 15.7a and 15.7b show the point spread functions for outcoupled TE and TM respectively. Outcoupled TE is nearly an Airy pattern although the exponential decay does broaden the distribution in one direction. Outcoupled TM is significantly degraded by the polarization variation across the pupil.

The Strehl ratio for the example given was about 0.1 for TM. Figs. 15.8a and 15.8b show the polarization in the far-field for outcoupled TE and outcoupled TM radiation. Note the significant variation in polarization in the side lobes for outcoupled TM radiation.

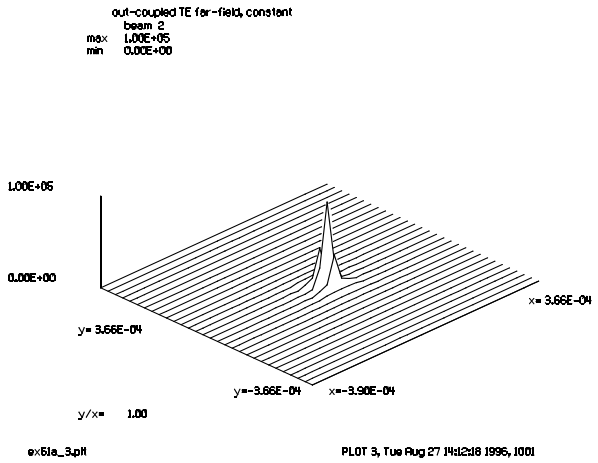


Fig. 15.7a. Far-field intensity of outcoupled TE radiation. The distribution is fairly close to an Airy pattern. The Strehl ratio is high.

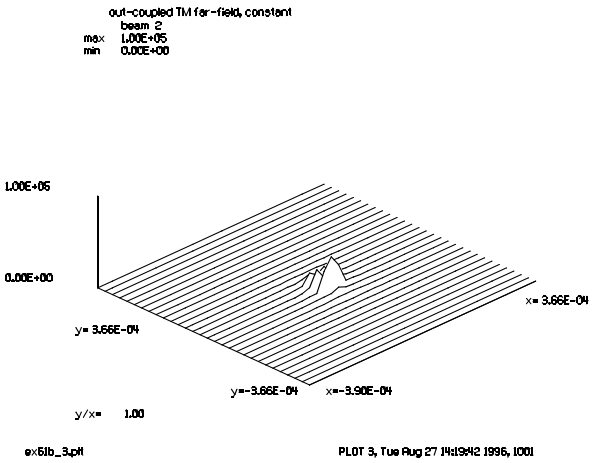


Fig. 15.7b. Far-field intensity of outcoupled TM radiation. The Strehl ratio is 0.1 because of the polarization variation across the pupil.

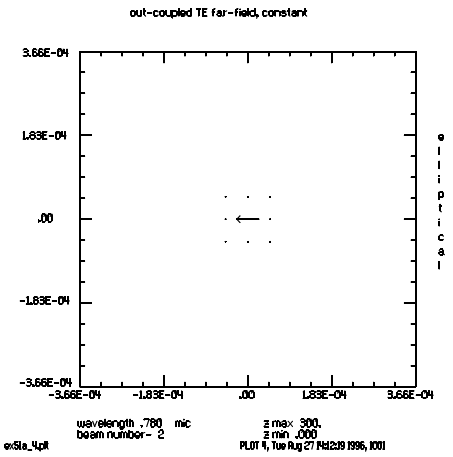


Fig. 15.8a. Far-field polarization map of outcoupled TE radiation.

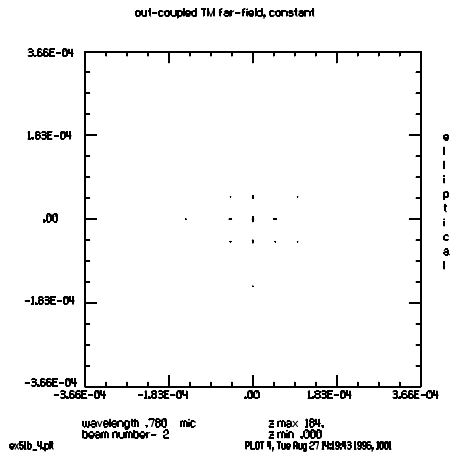


Fig. 15.8b. Far-field polarization map of outcoupled TM radiation. The central lobe shows orthogonal polarization in the central lobe of the outcoupled far-field TE plot. The Strehl ratio is less and information on the pupil polarization variation is contained in the secondary lobes.

15.3 Mode Coupling Theory

The outcoupled light may be characterized as a mode which has a characteristic complex amplitude and polarization function. For an ideal device the phase will yield a spherical wavefront. The intensity will vary according to an exponential decay. For simple collimated propagation under the grating, neglecting diffraction, and assuming an amplitude decay factor $\kappa(x, y)$, the amplitude varies as

$$E(x, y) = E(x, 0)e^{\int -\kappa(x, y)dy}, \tag{15.38}$$

where $E(x, 0)$ is the initial amplitude and y is the propagation direction. The integral in the exponent simplifies to y if the decay factor is constant.

The coupling constant varies according to the depth of the grating modulation and according to the direction of the coupled radiation. Coupled mode theory provides the basis for calculating the coupling coefficient.

$$\kappa = \kappa_0 \frac{\iint \boldsymbol{E}_g^{\dagger} \Delta\varepsilon \boldsymbol{E}_r dx dy}{\sqrt{\iint |\boldsymbol{E}_g|^2 dx dy} \sqrt{\iint |\boldsymbol{E}_r|^2 dx dy}}, \tag{15.39}$$

where the integration is taken transverse to the propagation direction. $\boldsymbol{E}_g$ is the electric vector mode for the guided field, $\boldsymbol{E}_r$ is the electric vector mode for the radiated field, † indicates the conjugate transpose vector, and $\Delta\varepsilon$ indicates the variation function of the electric permitivity. For an isotropic material $\Delta\varepsilon$ is a scalar. For anisotropic material $\Delta\varepsilon$ is a matrix. At this point, only isotropic material will be considered.

$\boldsymbol{E}_g$ may be written in terms of three orthogonal modes

$$E_g = \psi_x \hat{x} + \psi_y \hat{y} + \psi_z \hat{z}. \tag{15.40}$$

For TE and TM we have

$$E_{TE} \psi_x(z) \hat{x}, \tag{15.41}$$

$$E_{TM} = E_{TM_z} \phi(z) \hat{z} + E_{TM_y} \phi(z) \hat{y}, \tag{15.42}$$

where the ψ and ϕ functions are normalized eigenfunctions for TE and TM respectively.

We can break Eq. (15.39) into parts to determine the generation of s- and p-components. The three coefficients needed are

$$\kappa_{TE} = (\hat{\boldsymbol{e}}_{g_{TE}} \cdot \boldsymbol{e}_r) \iint \Delta\varepsilon(z) \psi_x(z) dx dz, \tag{15.43}$$

$$\kappa_{TM_z} = (\hat{\boldsymbol{e}}_{g_{TM_z}} \cdot \boldsymbol{e}_r) \iint \Delta\varepsilon(z) \psi_z(z) dx dz, \tag{15.44}$$

$$\kappa_{TM_y} = (\hat{\boldsymbol{e}}_{g_{TM_y}} \cdot \boldsymbol{e}_r) \iint \Delta\varepsilon(z) \psi_y(z) dx dz. \tag{15.45}$$

Since the radiation mode is a plane wave, it does not show up in the coupling coefficients except for the vector dot product. The vector dot products will be treated explicitly below. The coefficients of Eqs. (15.43) through (15.45) may be evaluated at only one point and only the geometric cosine factors calculated for each (x, y) point.

15.4 TE Guided Mode

The solution for the TE guided mode proceed as follows,

$$\text{TE guided field: } E_{TE}\psi_x(z)\hat{\boldsymbol{x}}. \tag{15.46}$$

The projection of $\hat{\boldsymbol{x}}$ onto the transverse plane of the radiation is

$$\tilde{\boldsymbol{x}} = \hat{\boldsymbol{x}} - (\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{k}})\hat{\boldsymbol{k}}. \tag{15.47}$$

The radiation vector is therefore

$$E_r = -j\sqrt{2\kappa_{TE}\langle z\rangle}E_x\tilde{\boldsymbol{x}}, \tag{15.48}$$

where $\langle z\rangle$ is the mean mode thickness.

Equation (15.48) defines the magnitude and polarization of radiation outcoupled from a TE guided mode. The radiation vector may be represented as components of $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{p}}$

The radiation will take the form

$$\boldsymbol{E}_r = E_s\hat{\boldsymbol{s}} + E_p\hat{\boldsymbol{p}}, \tag{15.49}$$

using the usual definitions of $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{p}}$ (See Eq. (11.38) and Eq. (11.39)).

$$E_s = \boldsymbol{E}_r \cdot \hat{\boldsymbol{s}},\ E_p = \boldsymbol{E}_r \cdot \hat{\boldsymbol{p}}. \tag{15.50}$$

The amplitude loss coefficient for the guided mode is

$$\kappa = \kappa_{TE}|\tilde{\boldsymbol{x}}|^2. \tag{15.51}$$

15.4.1 TM Guided Mode

The treatment of TM is similar but both z- and y-components must be treated

$$\text{TM guided mode: } E_{TM} = E_{TM_x}\varphi_x(z)\hat{\boldsymbol{z}} + E_{TM_y}\varphi_y(z)\hat{\boldsymbol{y}}, \tag{15.52}$$

$$\tilde{\boldsymbol{y}} = \hat{\boldsymbol{y}} - (\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{k}})\hat{\boldsymbol{k}}, \tag{15.53}$$

$$\tilde{\boldsymbol{z}} = \hat{\boldsymbol{z}} - (\hat{\boldsymbol{z}} \cdot \hat{\boldsymbol{k}})\hat{\boldsymbol{k}}. \tag{15.54}$$

The radiation is the sum of the z- and y-components

$$\boldsymbol{E}_r = -j\sqrt{2\kappa_{TM_z}\langle z\rangle}E_z\tilde{\boldsymbol{z}} - j\sqrt{2\kappa_{TM_y}\langle z\rangle}E_y\tilde{\boldsymbol{y}}. \tag{15.55}$$

Equation (15.55) defines the magnitude and polarization of radiation outcoupled from a TM guided mode. The loss coefficient for the guided mode is determined to conserve energy

$$\kappa = \kappa_0 \frac{\kappa_{TM_z}|E_z|^2|\tilde{\boldsymbol{z}}|^2 + \kappa_{TM_y}|E_y|^2|\tilde{\boldsymbol{y}}|^2}{|E_z|^2 + |E_y|^2}, \tag{15.56}$$

$$E_{TM}(y) = E_{TM}(0)e^{-\kappa y}. \tag{15.57}$$

15.4.2 Paraxial Approximation

For small numerical apertures and radiation exiting nearly normal to the surface

$$\boldsymbol{E}_r = E_s\hat{\boldsymbol{s}} + E_p\hat{\boldsymbol{p}} \approx E_x\hat{\boldsymbol{x}} + E_y\hat{\boldsymbol{y}}, \tag{15.58}$$

$$E_r = E_s(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{s}}) + E_p\sqrt{1-(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{s}})^2}, \tag{15.59}$$

$$E_r = -E_s\sqrt{1-(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{s}})^2} + E_p(\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{s}}). \tag{15.60}$$

The x- and y-components represent the properties of the exit pupil of the focused beam. The TE mode is established as a complex amplitude distribution with only an x-component. TM is established with y- and z-component. The z-component must be $\pi/2$ out of phase with respect to the x-component.

The electric vector components for TM are defined by Eqs. (15.36) and (15.37). The complex amplitude distribution is set up initially in the form

$$\begin{bmatrix} 0 \\ E_{TM_z} \\ E_{TM_y} \end{bmatrix} = E_{TM}\begin{bmatrix} 0 \\ \sqrt{1-c^2} \\ jc \end{bmatrix}, \tag{15.61}$$

where c is a purely real number such that the vector is normalized.

15.5 Bragg Condition

From Welford[3], we have the equations for solving the Bragg condition. Let the grating be defined by two ray vectors $\hat{\boldsymbol{r}}_{c_1}$ and $\hat{\boldsymbol{r}}_{c_2}$ that would constitute construction optics. Let the incident ray be defined by $\hat{\boldsymbol{r}}_{u_1}$ and the exiting ray be defined by $\hat{\boldsymbol{r}}_{u_2}$. The optical ray is the index times the unit direction vector. The Bragg condition is for ordinary free space optics is

$$\hat{\boldsymbol{n}} \times \left(\hat{\boldsymbol{r}}_{u_2} - \hat{\boldsymbol{r}}_{u_1} \right) = \frac{m\lambda_u}{\lambda_c} \hat{n} \times \left(\hat{\boldsymbol{r}}_{c_2} - \hat{\boldsymbol{r}}_{c_1} \right). \tag{15.62}$$

In the special case where beam "1" is in a waveguide and beam "2" is in air.

$$\hat{\boldsymbol{n}} \times \left(\frac{\hat{\boldsymbol{r}}_{u_2}}{\lambda_u} - \frac{n_{u_{eff}} \hat{\boldsymbol{r}}_{u_1}}{\lambda_u} \right) = m\hat{\boldsymbol{n}} \times \left(\frac{\hat{\boldsymbol{r}}_{c_2}}{\lambda_c} - \frac{n_{c_{eff}} \hat{r}_{c_1}}{\lambda_c} \right). \tag{15.63}$$

Define,

$$\boldsymbol{U} = \hat{\boldsymbol{n}} \times \frac{n_{u_{eff}} \hat{\boldsymbol{r}}_{u_1}}{\lambda_u} + m\hat{\boldsymbol{n}} \times \left(\frac{\hat{\boldsymbol{r}}_{c_2}}{\lambda_c} - \frac{\hat{\boldsymbol{r}}_{c_1}}{\lambda_c} \right), \tag{15.64}$$

such that

$$\hat{\boldsymbol{n}} \times \left(\hat{\boldsymbol{n}} \times \frac{\hat{\boldsymbol{r}}_{u_2}}{\lambda_u} \right) = \hat{\boldsymbol{n}} \times \boldsymbol{U}. \tag{15.65}$$

The solution is

$$\boldsymbol{k}_u = \boldsymbol{U} - (\hat{\boldsymbol{n}} \cdot \boldsymbol{U})\hat{\boldsymbol{n}} + \hat{\boldsymbol{n}}[1 - (\boldsymbol{U} \cdot \boldsymbol{U}) + (\hat{\boldsymbol{n}} \cdot \boldsymbol{U})^2]^{1/2}. \tag{15.66}$$

If the focus is at (x, y, z) and the point of diffraction on the grating is $(0, y, z)$

$$r_{c_2} = \frac{x_o \hat{\boldsymbol{x}} + (y_0 - y)\hat{\boldsymbol{y}} + (z_0 - z)\hat{\boldsymbol{z}}}{\sqrt{x_0^2 + (y_o - y)^2 + (z_0 - z)^2}}. \tag{15.67}$$

15.6 Round Trip: Outcoupling Followed by Incoupling

Fig. 15.9 illustrates out-coupling from a focusing waveguide coupler followed by reflection from a surface and subsequent in coupling. Figs. 15.10a to 15.10d illustrate the waveforms.

References

1. J. D. Jackson, C*lassical Electrodynamics*, Chap. 9, John Wiley & Sons, Inc. (1962).

2. A. Yariv, *Introduction to Optical Electronics*, 2nd Edition, Chap. 2, Holt, Rinehart and Winston (1971).

3. W. T. Welford, *Aberrations of Optical Systems*, Section 5.4 and 5.5, Adam Hilger, (1986).

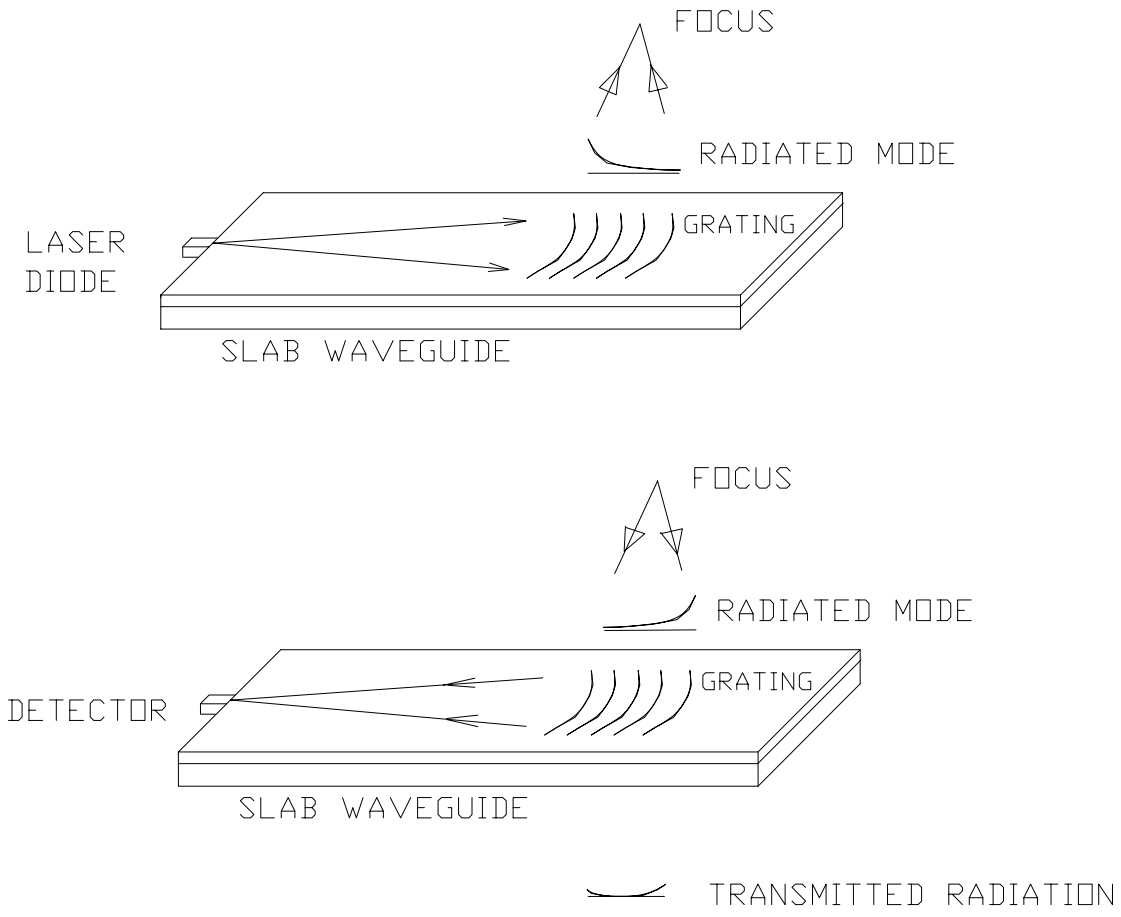


Fig. 15.9. Schematic illustrating outcoupling of radiation, focusing on an optical data storage surface, reflection, and return of the radiation with the distribution reversed. In outcoupling (top), light from a laser diode is injected into a slab waveguide. The light propagates as a guided mode in the vertical direction — trapped in a boundary layer of higher index. A grating of variable index or surface relief is put on the waveguide grating. The grating lines are curved to form a diffracted beam which converges to the focus. The guided mode decays as it passes under the grating, as the light is scattered out. The beam is reflected at the focus by the optical data storage surface and returns to the grating (bottom). The returned, incident mode is flipped with respect to the outgoing beam and is, therefore, not mode matched. The light is partly scattered into the grating and partly transmitted, because the mode is not perfectly matched. The guided mode is focused onto a detector. A more detailed arrangement —not shown for simplicity — allows optical separation of the laser diode and the detector paths.

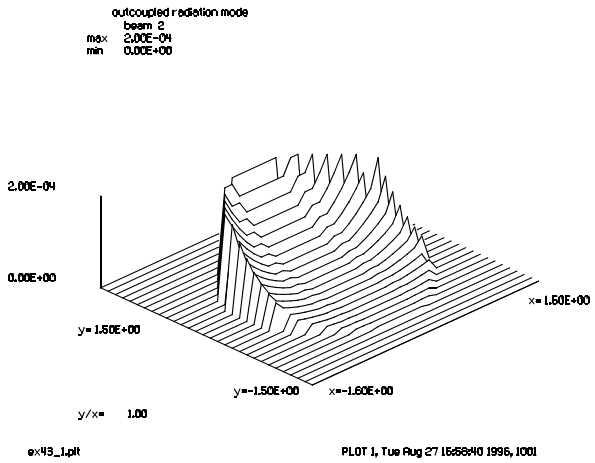


Fig. 15.10a. Radiated mode after outcoupling.

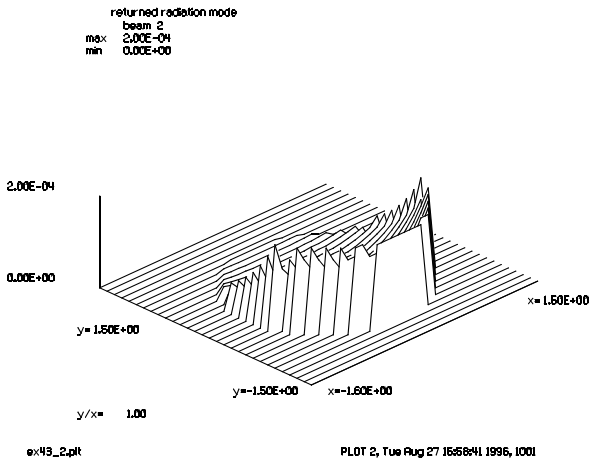


Fig. 15.10b. Radiated mode after outcoupling.

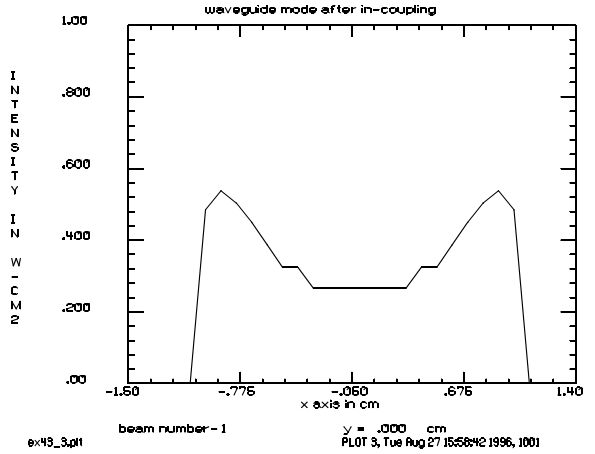


Fig. 15.10c. Radiated mode after outcoupling.

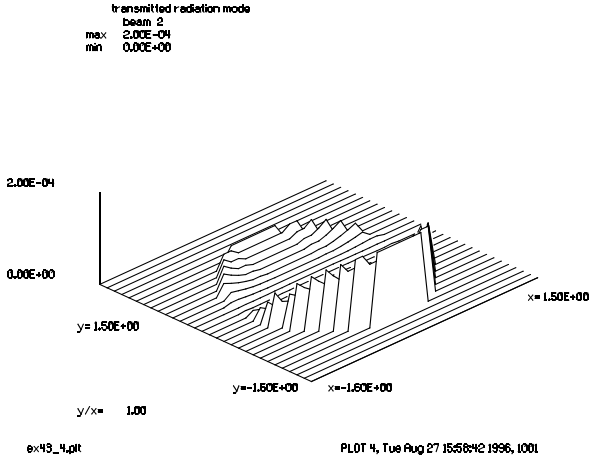


Fig. 15.10d. Transmitted radiation mode. Note that the radiation is most poorly matched at the upstream and down stream parts of the grating aperture.

16. Optimization

It is a great advantage to be able to have automatic optimization of the parameters of the system. This can be done using the registers for variables, targets, and constraints. The variables may be any real number since any real number can be controlled by registers. The target values can be any of the parameters accessible from the command `variable/set/parameter`. The optimization is implemented by using macros. Figure 16.1 illustrates the operations. One begins by initializing the problem and then call the optimization macro. If the optimization is successful stop, if not a second call to the optimization macro is made.

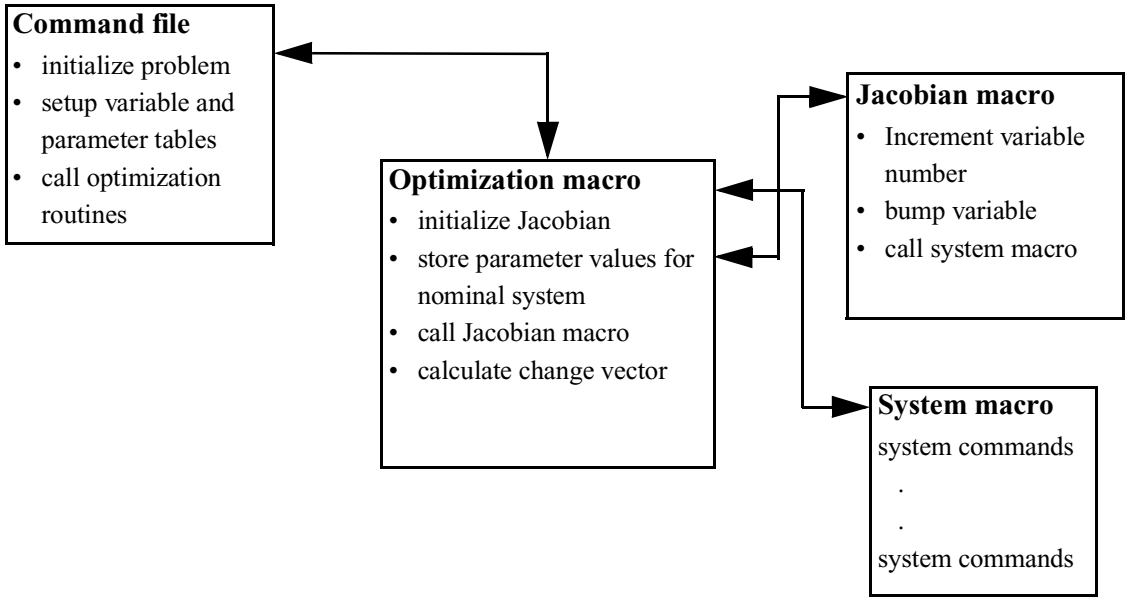


Fig. 16.1. Schematic of command and macros for optimization.

The optimization macro initializes the nominal target values by calling the Jacobian matrix with the nominal variable values. The optimization macro then calls the Jacobian macro to calculate the Jacobian matrix and ultimately to calculate the change vector. The Jacobian macro bumps each variable in turn and calculates the value of the target parameters by calling the system macro. The system macro responds to the value of the variables and calculates the parameter values. When the Jacobian matrix calculates the change table is generated and the corrected set of variable values is created. The command sequence sets the size of the array, identifies the name of the system macro (in this case called `system`), defines the variables and parameters, and initializes the system. For this example spherical aberration and coma are added. The commands to set up a sample optimization are listed in Fig. 16.2. The OPT macro, listed in Fig. 16.3, initializes the Jacobian by calling the system macro with the nominal variable values. The Jacobian macro is called twice — once for each variable. The Jacobian macro increments the variable number and bumps

the current variable, calls the system macro, and stores calculated parameter values. Upon return to the optimization macro OPT, the Jacobian matrix is calculated and listed and the change table can be calculated.

Fig. 16.2. Sample command sequence for optimization.

```
echo/on                              # set echo on for commands
array/s 1 16                         # define array size
nbeam 2                              # define two beams
opt/nam system                       # define macro name, this feature is not used now
opt/var/add Focus .01                # first variable
opt/var/add Tilt .01                 # second variable
opt/tar/add Str 1                    # Strehl with target 1
clap/c/c 1 5                         # initialize problem
copy 1 2
abr/sph 2 .2                         # define aberrations
abr/coma 2 .2
macro/run opt                        # run optimization macro once
```

Fig. 16.3. Macro commands for optimization and system functions.

```
.opt
 opt/jac/run 1                       # 1 = initialize Jacobian
 macro system                        # run system macro
 opt/jac/run 2                       # 2 = store nominal performance values
 macro jacobian/2                    # call jacobian macro once for each variable
 opt/jac/list                        # list Jacobian values, all done with Jacobian
 opt/change                          # compute and implement change table.
macro/end
.jacobian
 opt/jac/run 3                       # 3 = increment each variable
 Str=                                # list value of Strehl ratio
 macro system                        # run system macro
 opt/jac/run 4                       # 4 = store all target values
macro/end
.system
 copy 2 1                            # system macro, copy stored beam to beam 1
 abr/focus 1 Focus                   # add some defocus through register 1
 abr/tilt 1 Tilt                     # add some tilt through register 2
 variab/set/par Str 1 strehl         # store Strehl ratio in register 3 as parameter.
macro/end
```

16.1 Method of Least Squares

The optimization process consists of manipulating a set of parameters $x_1, x_2, \ldots x_n$ to minimize a merit function ϕ defined in terms of a set of performance parameters $f_1, f_2, \ldots f_n$ where f_i are weighted aberrations of the optical system. The vector of performance parameters is

$$f = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \cdot \end{bmatrix}, \tag{16.1}$$

where

$$f_i = w_i(e_i - t_i), \tag{16.2}$$

w_i is the weight, e_i is the current value, and t_i is the target value. The aberrations are functions of the system variables, x_i, such that (to first order)

$$\phi(x) = f_1^2(x) + f_2^2(x) + \ldots f_m^2(x) = \boldsymbol{f}^t\boldsymbol{f}. \tag{16.3}$$

where the superscript *t* indicates the transpose. In vector notation,

$$\phi = \boldsymbol{f}^t\boldsymbol{f}. \tag{16.4}$$

We define the Jacobian to be

$$\mathbf{A} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdot & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdot & \dfrac{\partial f_2}{\partial x_n} \\ \cdot & \cdot & \cdot & \cdot \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdot & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}. \tag{16.5}$$

The aberrations are functions of the system variables, x_i, such that (to first order)

$$\boldsymbol{f} = \mathbf{A}\boldsymbol{x}. \tag{16.6}$$

The matrix **A** is the Jacobian defined such that

$$a_{ij} = \frac{\partial f_i}{\partial x_j}. \tag{16.7}$$

Given an initial aberration vector f_o for the variable state x_o, the first order relationship is

$$\boldsymbol{f} = \boldsymbol{f}_0 + \mathbf{A}(\boldsymbol{x} - \boldsymbol{x}_0). \tag{16.8}$$

The optimization procedure is to find the vector $\boldsymbol{x}$ which minimizes ϕ,

$$\phi = (\boldsymbol{f}_0 + \mathrm{A}(\boldsymbol{x} - \boldsymbol{x}_0))^t(\boldsymbol{f}_0 + \mathrm{A}(\boldsymbol{x} - \boldsymbol{x}_0)). \tag{16.9}$$

The derivative with respect to the vector of variables is

$$\frac{d\phi}{d\boldsymbol{x}} = 2\mathbf{A}^t\mathbf{A}(\boldsymbol{x} - \boldsymbol{x}_0) + 2\mathbf{A}^t\boldsymbol{f}_0. \tag{16.10}$$

Setting $\frac{d\phi}{d\boldsymbol{x}} = 0$ and solving, we have

$$\mathbf{A}^t\mathbf{A}(\boldsymbol{x}-\boldsymbol{x}_0) = -\mathbf{A}^t\boldsymbol{f}_0, \tag{16.11}$$

with solution

$$\boldsymbol{x}-\boldsymbol{x}_0 = -(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t\boldsymbol{f}_0. \tag{16.12}$$

This is the method of least squares.

16.1.1 Damping

Because the DLS method described above is a first order approximation and therefore neglects nonlinearities of the system, it is prone to oscillations. In order to add stability to the procedure we add into the merit function the square of the change vector times a damping factor,

$$\phi \rightarrow \phi + p^2(\boldsymbol{x}-\boldsymbol{x}_0)^t(\boldsymbol{x}-\boldsymbol{x}_0). \tag{16.13}$$

where p^2 is the damping factor. Strong nonlinearities in the system require larger values of p^2, i.e., more damping. Solving for the modified merit function, we have

$$\boldsymbol{x}-\boldsymbol{x}_0 = -(\mathbf{A}^t\mathbf{A}+p^2\mathbf{I})^{-1}\mathbf{A}^t\boldsymbol{f}_0, \tag{16.14}$$

where **I** is the unity matrix.

Having taken at least one iteration, we have calculated a set of aberrations $\boldsymbol{f}_{1c}$ which we can compare with the actual aberrations $\boldsymbol{f}_1$ at the state $\boldsymbol{x}_1$, so we can determine the residual aberrations. It is possible to solve for the value of p^2 by recalculating $\boldsymbol{f}_{1c}$. The optimum value of damping factor may be used for the next optimization cycle by minimizing the difference $\boldsymbol{f}_1 - \boldsymbol{f}_{1c}$ and using that damping factor for the next iteration.

16.1.2 Applying Constraints

It is commonly required that the optical system meet certain constraints, i.e., focal length, overall length, positive edge thicknesses, etc. Constraints may be equalities or inequalities. In the simplest approach inequalities are ignored completely if the system does not violate them and applied as equalities if the system does violate them. In practice this may result in oscillation at a constraint boundary and soft boundaries may be required. The discussion here is limited to equalities since this topic is fundamental.

Consider a set of constraint functions $q_k(\boldsymbol{x})$ and a set of target values c_k. For example we may wish to control the overall power in terms of the curvatures of two thin lenses in proximity,

$$(n_1-1)(C_{11}-C_{12})+(n_2-1)(C_{21}-C_{22}) = K, \tag{16.15}$$

$$q_1(C_{11}, C_{12}, C_{21}, C_{22}) = (n_1 - 1)(C_{11} - C_{12}) + (n_2 - 1)(C_{21} - C_{22}), \tag{16.16}$$

$$x_1 = C_{11}, x_2 = C_{12}, x_3 = C_{21}, \text{and } x_4 = C_{22}. \tag{16.17}$$

16.1.3 Lagrange Multipliers

If we approximate the functions q_k to first order, then we have the equation

$$\mathbf{B}(\boldsymbol{x} - \boldsymbol{x}_0) = \boldsymbol{c} - \boldsymbol{c}_0, \tag{16.18}$$

where $\boldsymbol{c}$ is the vector of target constraint properties, $\boldsymbol{c}_0$ is the vector of initial constraints, and **B** is the Jacobian matrix consisting of elements

$$b_{ij} = \frac{\partial q_i}{\partial x_j}. \tag{16.19}$$

Thus the procedure for finding the solution vector to set the constraints properly is identical to the DLS procedure for finding the aberrations. However we wish to maintain the constraints once having met them. It is possible to maximally control the constraint conditions by restricting parameter state movements to directions which are normal to the surfaces of constant c_k. The row vectors of **B** define directions normal to the constraint surfaces. For example,

$$\begin{bmatrix} b_{i1} & b_{i2} & \cdots \end{bmatrix} = \begin{bmatrix} \frac{\partial q_i}{\partial x_1} & \frac{\partial q_i}{\partial x_2} & \cdots \end{bmatrix}. \tag{16.20}$$

The possible directions are specified by $\mathbf{B}\Delta\boldsymbol{\lambda}$ where $\Delta\boldsymbol{\lambda}$ is a vector of coefficients defining the magnitude of movement along the various constraint directions.

From the definition of the merit function

$$\phi = \boldsymbol{f}^t\boldsymbol{f}, \tag{16.21}$$

and with $\boldsymbol{f} = \mathbf{A}\boldsymbol{x}$ and $\nabla\phi = 2\mathbf{A}^t\boldsymbol{f}$.

We can solve for the vector of Lagrange coefficients to achieve

$$\nabla\phi = 2\mathbf{B}^t\Delta\boldsymbol{\lambda}. \tag{16.22}$$

This results in the matrix equation

$$\begin{bmatrix} \mathbf{A}^t\mathbf{A} & \mathbf{B}^t \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\boldsymbol{x} \\ \Delta\boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\mathbf{A}^t\boldsymbol{f}_0 \\ -\Delta\boldsymbol{c} \end{bmatrix}. \tag{16.23}$$

Solution of the augmented matrix equation will result in close adherence to the required constraints. The set of Lagrange multipliers, $\Delta\boldsymbol{\lambda}$, are not explicitly used.

The Lagrange multiplier method may be interpreted as two effects. The first effect is the resolution of the required change in constraint values which could be taken as an extension of the previous least squares solution to include the constraint equation,

$$\mathbf{A}^t\mathbf{A}(\boldsymbol{x}-\boldsymbol{x}_0) = -\mathbf{A}^t\boldsymbol{f}_0 \text{ and } \mathbf{B}(\boldsymbol{x}-\boldsymbol{x}_0) = \boldsymbol{c}-\boldsymbol{c}_0, \tag{16.24}$$

$$\begin{bmatrix}\mathbf{A}^t\mathbf{A}\\ \mathbf{B}\end{bmatrix}\Delta\boldsymbol{x} = \begin{bmatrix}-\mathbf{A}^t\boldsymbol{f}_0\\ -\Delta\boldsymbol{c}\end{bmatrix}. \tag{16.25}$$

The second effect is ensuring that the solution moves parallel to the constraint surfaces

$$\mathbf{A}^t\mathbf{A}\Delta\boldsymbol{x} + \mathbf{B}^t\Delta\boldsymbol{\lambda} = -\mathbf{A}^t\boldsymbol{f}_0. \tag{16.26}$$

The term $\mathbf{B}^t\Delta\boldsymbol{\lambda}$ is in a direction perpendicular to the surfaces defining the constraints. Figure 16.4 illustrated the several vectors associated with Lagrange multipliers.

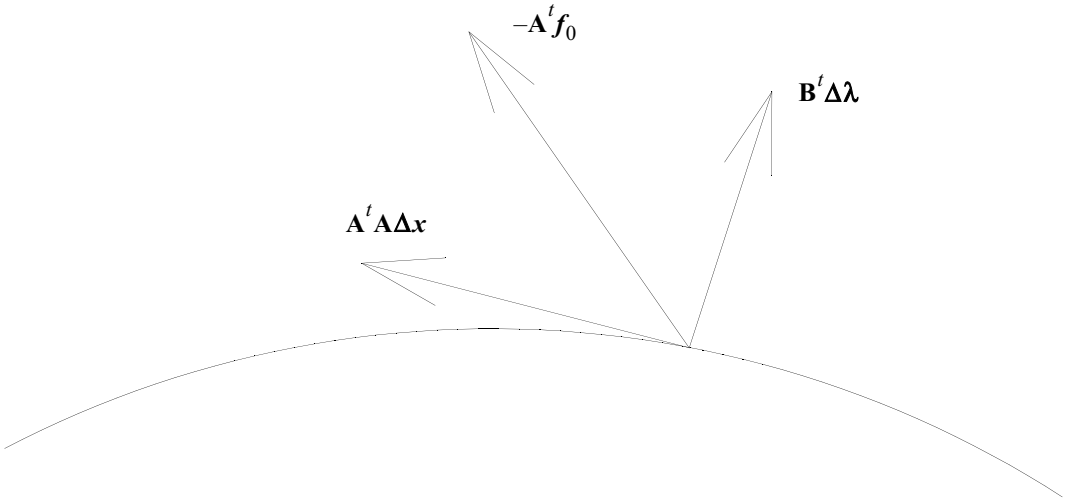


Fig. 16.4. Illustration of solving for vectors perpendicular and tangent to the constraint surface to fit the desired change in parameter space. We discard the vector $\mathbf{B}^t\Delta\boldsymbol{\lambda}$, keeping the solution on the constraint surface.

By fitting the two vectors on the right hand side of Eq. (16.26), we have terms perpendicular and tangent to the surfaces defining the constraints. We discard the perpendicular term assuring the solution moves tangent to the constraint surface.

Equation (16.23) is suitable for solving equality constraints. An example of an equality constraint is maintaining the focal length. An example of an inequality constraint is to keep the focal length greater or less than some value. Inequality constraints may be implemented by invoking Eq. (16.23) only when the

constraint is violated and repeating the optimization step. When the constraint is not violated we rewrite **B** to delete the row corresponding to the deleted constraint.

16.2 Example of Optimization

This example illustrates optimization in a very simple configuration. The GLAD/OPTION feature must be implemented. The system to be optimized is a spatial filter, as illustrated in Fig. 16.5. This problem is solved in Example 60. There is a pinhole at the focal region of elliptical shape with the x- and y-radii variables. The location is also allowed to vary. Astigmatism is introduced at the first lens and the objective of the optimization is to optimize simultaneously the Strehl ratio and the energy. The aperture will be moved and resized by the optimization process to roughly match the line foci caused by the astigmatism. The merit function is

$$\phi = (\text{SR} - 1)^2 + (\text{Energy} - 1)^2 . \tag{16.27}$$

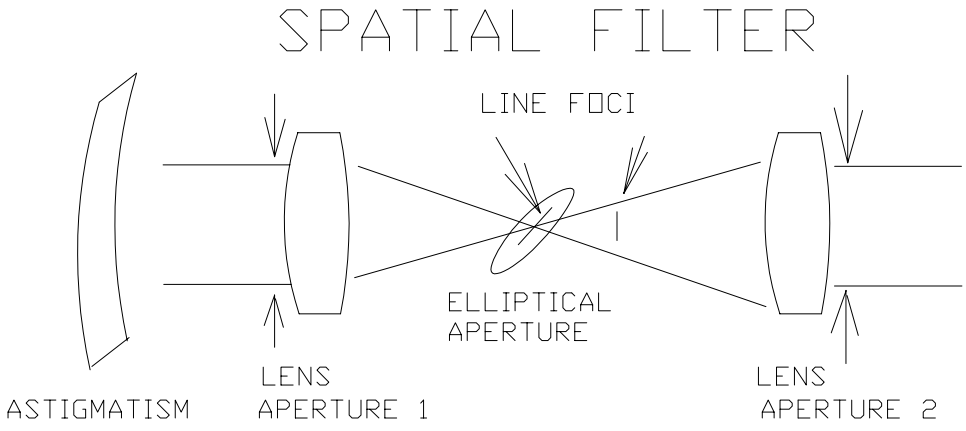


Fig. 16.5. Configuration to be optimized. The pinhole at the line focus is to be moved and its size changed in the x- and y-directions to simultaneously optimize the Strehl ratio and energy transmitted through Aperture 2.

In this particular case there are two solutions — the two line foci. Figures 16.6a, 16.6b, and 16.6c show the line foci to either side and the bilaterally symmetrical pattern at the paraxial focus. Fig. 16.6d shows the convergence is rapid for this simple problem. Convergence in more complex problems may not be as easy to achieve.

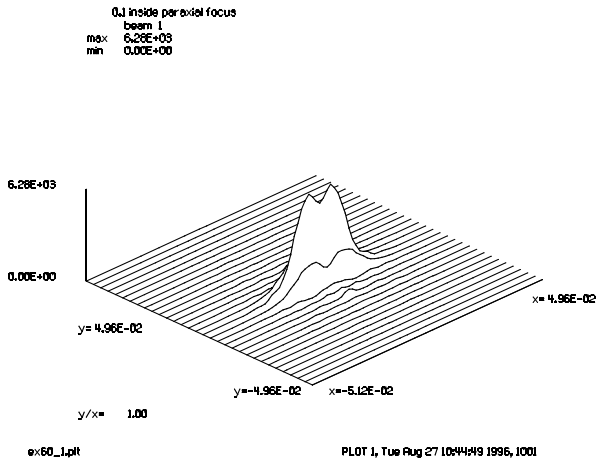


Fig. 16.6a. Line focus 0.1 inside the paraxial focus.

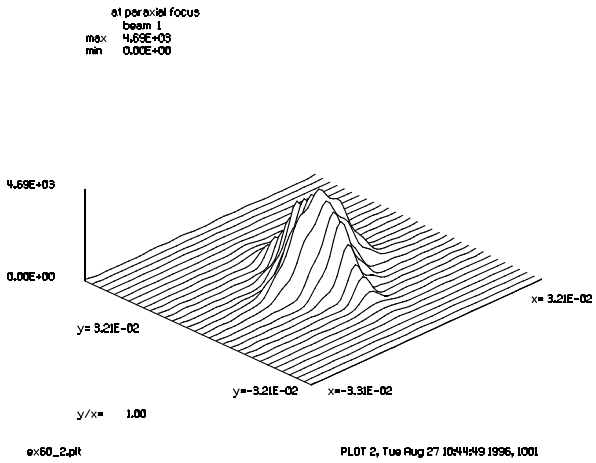


Fig. 16.6b. Paraxial focus.

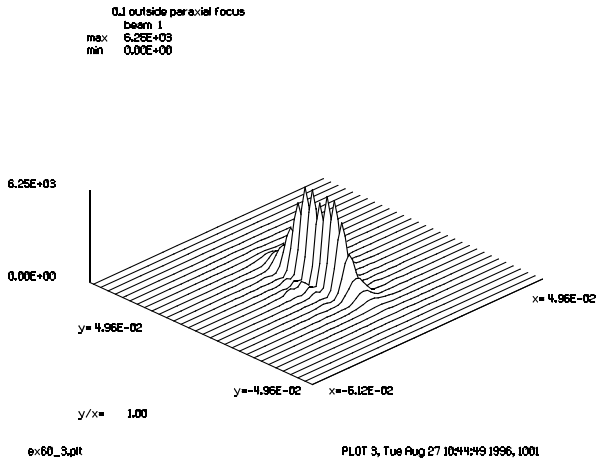


Fig. 16.6c. Line focus 0.1 outside the paraxial focus.

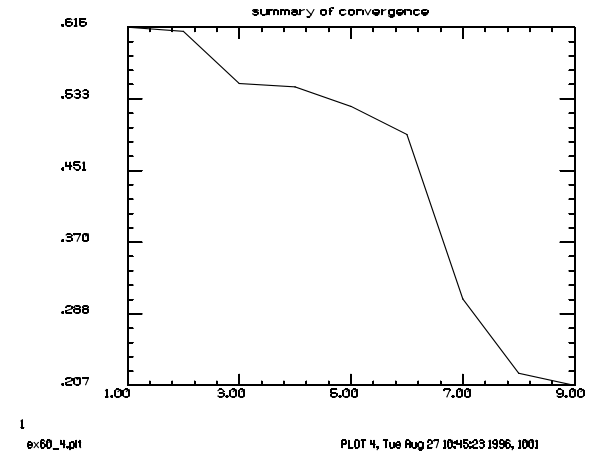


Fig. 16.6d. Merit function change per optimization cycle.

17. Axicon Mirrors

Most laser beams may be considered to be propagating along a well defined optical axis. In this chapter, some important exceptions will be considered. Cone mirrors, called axicon mirrors, can transform a conventional beam into a cylindrical wavefront which propagates in the radial direction. Axicons may be general surfaces of rotation but the scope here will be limited to the elementary conical form for simplicity. We shall refer to ordinary beams with a well defined axis as axial beams and cylindrical wavefronts as radial beams. Fig 17.1 illustrates the transformation of axial to radial mode and illustrates the fact that the optical axis becomes a plane in radial mode located at the apex of the axicon. Propagation of both axial and radial beams are readily treated numerically by Fourier methods. When considering axicon mirrors, it is convenient to generalize the concept of axis to include planar optical axes and cylindrical optical axes. The behavior of the optical axis is a key consideration in characterizing the properties of axicon optical systems.

A radial beam may be transformed back into an axial beam by a second axicon, but in the more general case the radial beam will be transformed into an annular form, as shown in Fig. 17.2. The annular beam is characterized by a cylindrical optical axis. In the special case where the radius of the annular axis goes to zero, the annular beam becomes identical to an axial beam.

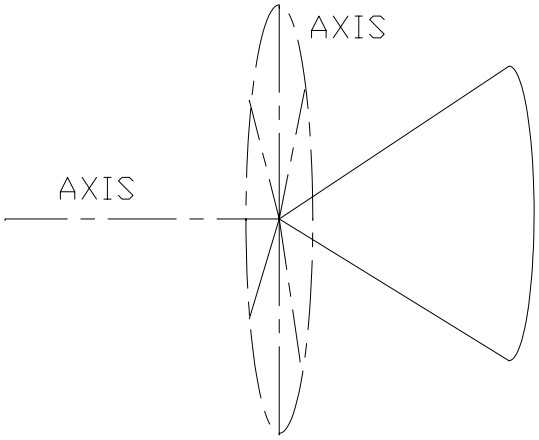


Fig. 17.1. An axicon mirror splits the optical axis of an axial beam into a plane perpendicular to the axis in forming a radial beam.

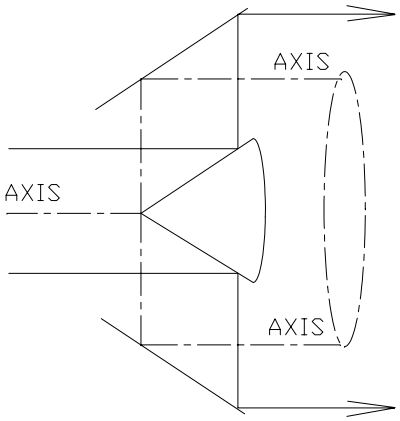


Fig. 17.2. A second axicon can generate an annular beam from a radial beam with the axis taking the form of a cylinder.

The `axicon` command controls a number of operations associated with axicons. GLAD is limited to the special case of axicons consisting of simple cone mirrors with 90º apex angles. Normally beams propagate as “pencils” of light with a well defined optical axis. This condition is referred to as axial mode. Axicon mirrors can transform axial beams into radially (or nearly radially) propagating beams. After an axial beam hits an axicon, the optical axis is split into a plane perpendicular to the original axis. This condition is referred to as radial mode. In the general case, an axicon will operate on a beam in radial mode to produce a beam with the optical axis transformed into a cylinder. This is referred to as annular mode. An

important special case is when the radius of the cylindrical axis of the annular beam degenerates into the original optical axis. This case constitutes a return to axial mode.

An important consideration in axial beams is the proper treatment of divergence. Spherical wavefronts may be as accurately treated as collimated wavefronts. Generally annular beams have been treated as semi-collimated beams. One may generalize the annular beam to treat noncollimated annular beams. The fast-Hankel transform developed by Siegman is useful only for semi-collimated annular beams[1–4]. We shall present a superior propagation method of propagating annular beams which circumvents the numerical difficulties of the fast-Hankel transform and which can treat beams with significant divergence or convergence.

17.1 Radial Mode Propagation

In collimated space, we represent the complex amplitude by an ensemble of plane waves. A cylindrical wave should be represented by a set of circumferential harmonic functions. The period of the harmonic functions should be integer fractions of the circumference. The intrinsic aliasing behavior of discrete FFT's works to our advantage in this application. Discrete one-dimensional Fourier transforms are identical in behavior to an infinite field of identical arrays. This is the same solution as the cylindrical harmonic functions provided we wrap the circumferential distribution into the rectangular array with no guard band in the circumferential direction — vertical axis in Fig. 17.3.

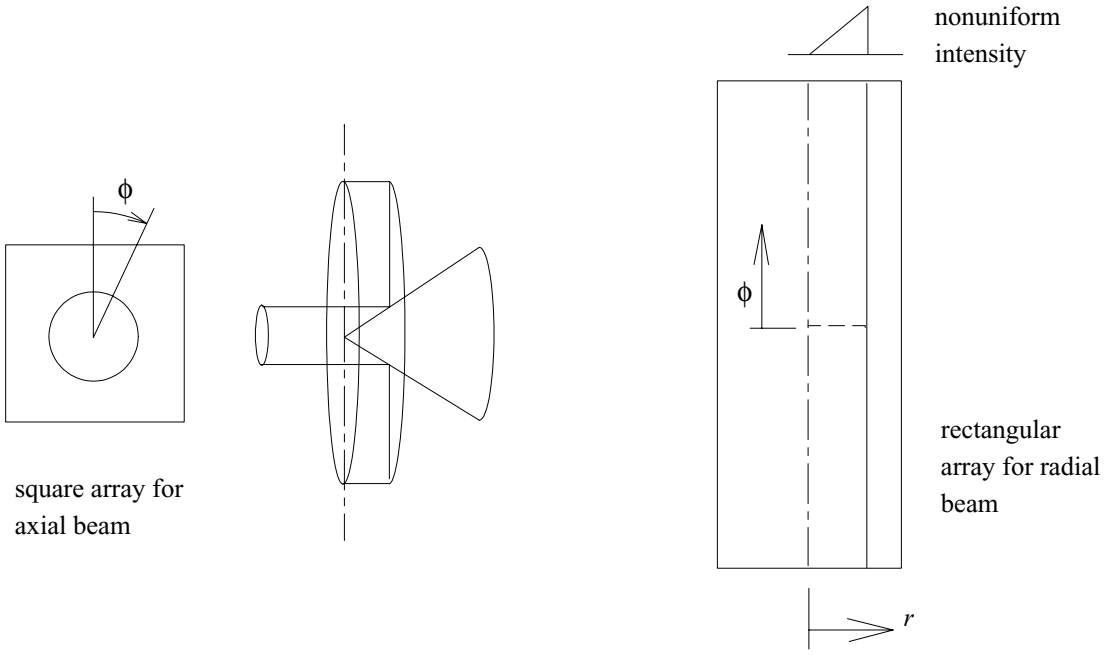


Fig. 17.3. Schematic of transformation of a flat top axial beam represented on a square array, into a radial beam with nonuniform intensity on a rectangular array. In radial mode, the horizontal axis represents the radius and the vertical axis represents the azimuthal direction. The optical axis is a vertical line in the center of the rectangular array. By using no guardband in the vertical direction, cylindrical propagation is treated correctly by FFT techniques.

Taking the circumferential direction to be the y-direction, the cylindrical propagator is,

$$a(y, R_2) = \left(\frac{R_1}{R_2}\right)^{1/2} \mathbf{F}^{-1}\left[e^{-j\pi z_{eff}\eta^2} \mathbf{F}[a(y, R_1]\right], \tag{17.1}$$

where $z_{eff} = (R_2 - R_1)(R_1/R_2)$, R_1 and R_2 are the radii of the cylindrical wave before and after the propagation, $\mathbf{F}$ and $\mathbf{F}^{-1}$ are the forward and inverse Fourier transforms, and η is the spatial frequency variable for the y-direction. GLAD allows separable diffraction propagation, but the existing routines must be modified for this application.

The x-direction may be treated in more conventional fashion. GLAD currently allows cylindrical optics and separable diffraction. This treatment is consistent with the use of a cylindrical nonlinear gain region. The radial beam may be represented numerically by a rectangular array with the x-direction corresponding to the radial direction and the y-direction corresponding to the azimuthal direction. Unlike axial beams, the azimuthal direction is arranged to exactly fill the numerical array. In this case the harmonic treatment intrinsic to Fourier transform methods accurately represents the cyclical nature of the cylindrical wave propagation. The transformation from axial form to radial form requires the use of interpolation. The inner radial zones in the axial representation are significantly stretched to fit into the radial representation, as shown in Fig. 17.4. The center point of the axial beam corresponds to the center of the x-axis in radial form. The center point is infinitely stretched out and must have zero intensity.

Given a uniform intensity in the initial axial beam, the resulting radial beam will vary linearly with radius. Figs. 17.5–17.8 show how an initial flat top function is transformed into a radial beam with linear variation of intensity. GLAD uses linearly interpolation which results in some raggedness of the boundary. The conversion of axial-to-radial modes generally shows loss in energy of 1% to 2%. This energy loss does not significantly effect propagation properties or the formation of laser modes.

It will be apparent that Eq. (17.1) will fail if we attempt to propagate to $R_2 \to 0$. This is to be expected because the cylindrical wave would, according to the simple Fresnel theory, have infinite image intensity. Clearly the Fresnel theory is unable to treat the cylindrical wave near the focus. Unlike axial beams where zones inside and outside the Rayleigh range are defined, the waist of the cylindrical wave is set to an infinitesimal value, the Rayleigh range is essentially zero, and the beam is always outside the Rayleigh range. The logical coordinate system for the cylindrical wave is to use the original optical axis as the center of the cylindrical wave calculations.

17.2 Diffraction Considerations in Generating a Radial Beam by an Axicon

The axicon is a “thick” element. The center of the beam spends more time in axial mode and less in radial mode than the outer part of the beam. This behavior creates a variation of the point spread function across the axis and, therefore, can not be represented by Fourier methods. Fig. 17.4 illustrates schematically the variation of the diffraction point spread function across the pupil. In the radial direction, the point spread function has the same width across the pupil. It should also be noted that the width of the diffraction point spread function is greatest in the radial direction, implying diffraction will be of most consequence in that direction. In the azimuthal direction, the beam is strongly diverging so that a given physical distance along the radius has a small effective diffraction length. In the azimuthal direction, the region near the cone tip begins to expand sooner than the outer regions of the beam, so the inner regions will show smaller diffraction effects and, therefore, the diffraction spread function is narrower in the azimuthal direction. The non-Fourier effects in the azimuthal direction are of relatively minor consequence because the diffraction in the radial

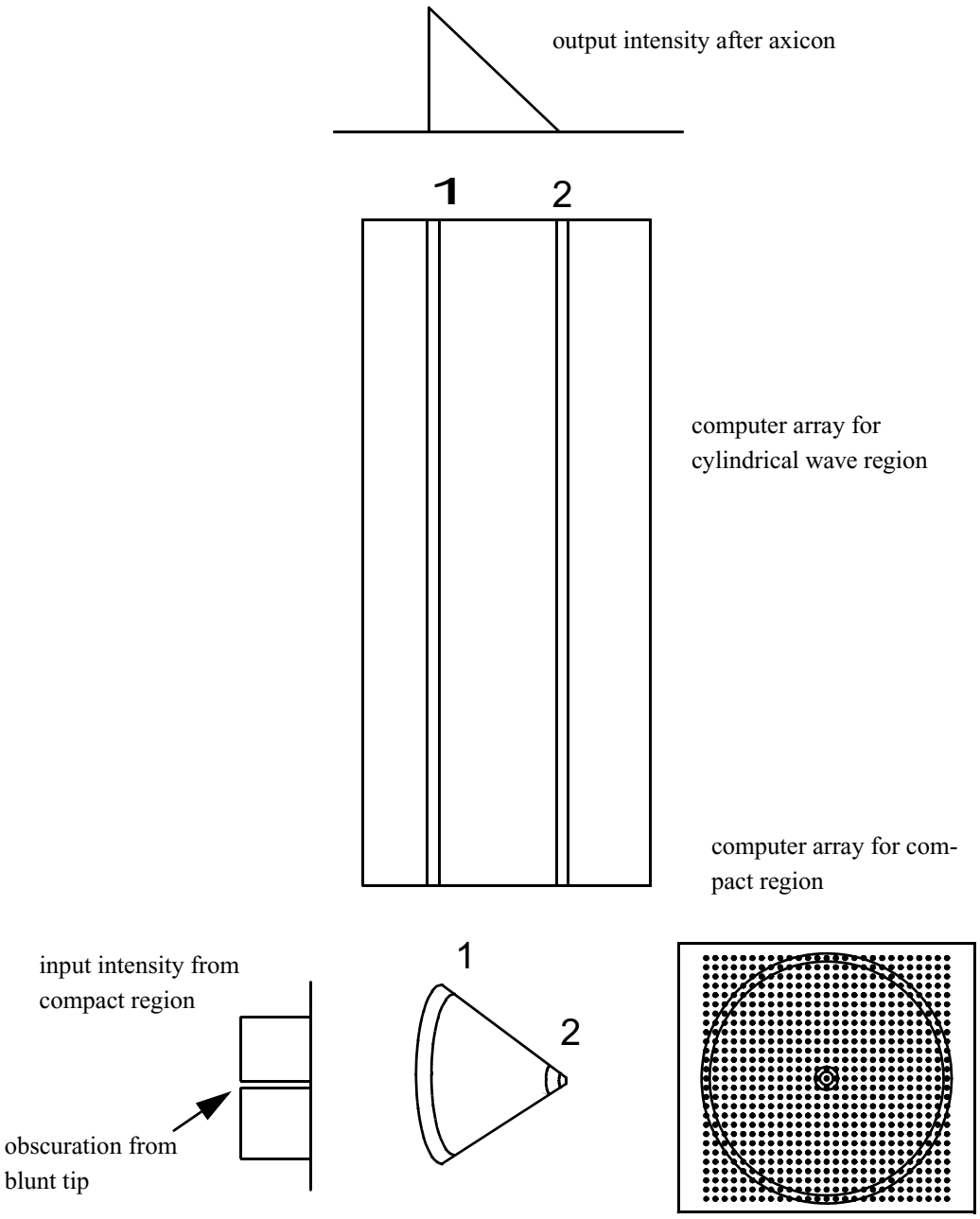


Fig. 17.4. The axicon function will map zones in the compact beam path into the radial wave representation and vice versa. The compact beam may be represented by a conventional square array. The inner zone 2 must be greatly stretched, reducing the intensity substantially. The tip of the zone can not be treated.

direction has more effect and most axial devices are nearly radially symmetric. If required, one could address the non-Fourier behavior by the aperture division methods described in Ex. 35.

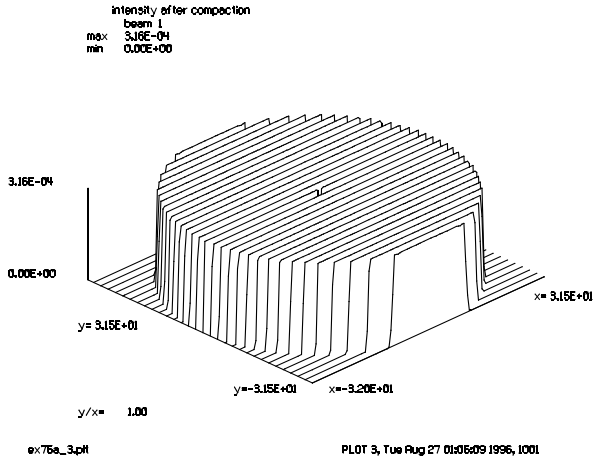


Fig. 17.5. Intensity in axial mode. The distribution has a hole in the center, representing the blunt tip of the cone.

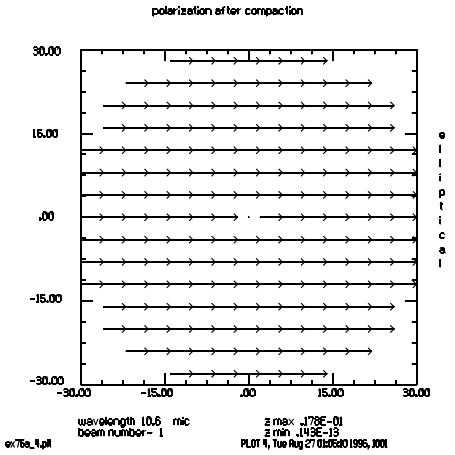


Fig. 17.6. The polarization state in the axial leg is linearly polarized.

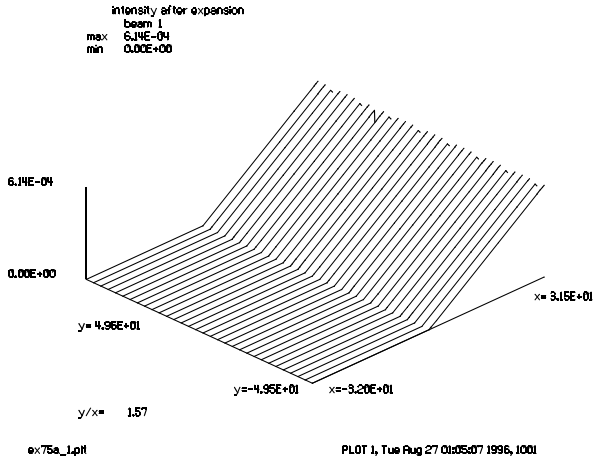


Fig. 17.7. Intensity distribution in radial beam showing the nonlinear effect on intensity. The radius is in the x-direction, the azimuth in y.

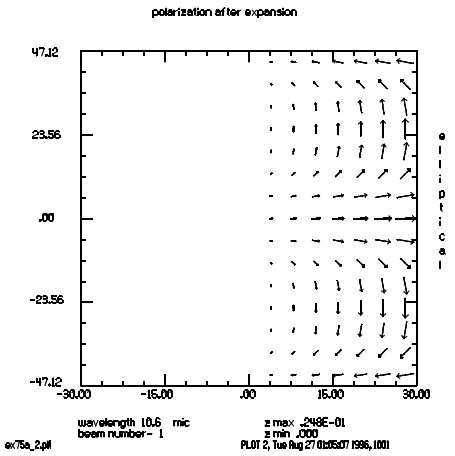


Fig. 17.8. Polarization in the radial beam. Note that the polarization rotates in the azimuthal direction.

The `axicon/radial` command uses a user-specified radius to determine the exiting radius of the radial beam. GLAD propagates that distance from the cone tip as an axial beam and then does a geometric conversion of the intensity, as illustrated in Fig. 17.9.

Axicons require special consideration of nonuniform irradiance and polarization disruption. The axicon remaps the irradiance in nonuniform fashion. From geometrical theory, intensity in each zone is changed by the ratio of the radius of that zone before and after the axicon. In propagating from an axial beam to a radially expanding beam, all axial zones are converted to zones in the radial beam with identical radii. This causes inner zones in the axial beam to have greatly reduced intensity in the radial beam. The process is reversed in converting radial beams to axial beams. Zones in the radial beam which are converging to the axicon tip have high intensity in the axial beam. In principle the intensity of the axial beam goes to infinity at the axicon tip after conversion from a radial beam. In reality, diffraction will cause the “hot tip” effect to spread. In converting from axial to radial mode or vice versa, GLAD requests specification of a finite radius of the radial beam. For axial-to-radial conversion the final state in radial mode is set at the specified radius. For radial-to-axial conversion, the beam is propagated in radial mode to the specified radius and then converted.

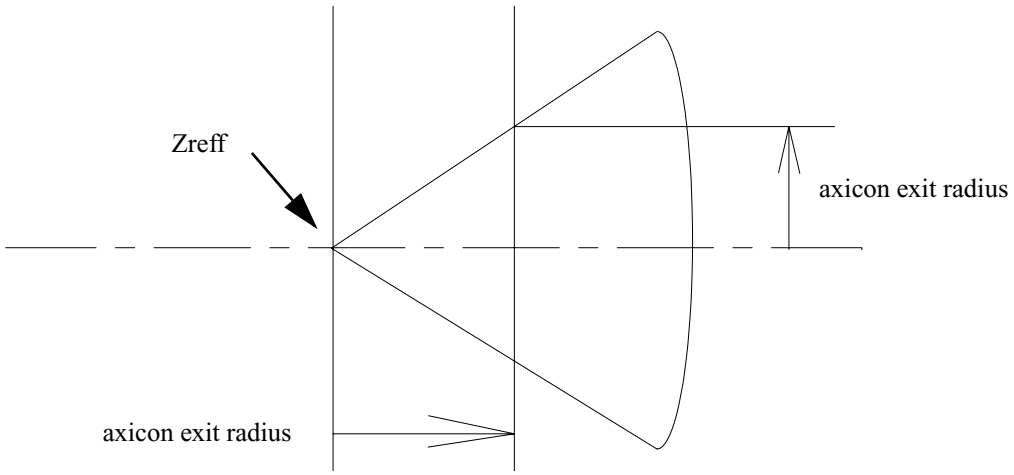


Fig. 17.9. A radius is specified for the exiting radius of the radial beam. GLAD propagates the distance past the cone tip and then does a geometric transformation to the radial mode.

With a radial beam generated with the `focal` parameter, toroidal mirrors may be used to control the degree of divergence as projected onto the optical axis. In the azimuthal direction the toroidal mirror must have the radius of the radial wavefront. The `mirror/toroid` command may be used with both radii specified correctly. Alternately, the parameter `xonly` may be used in the form `mirror/xcyl kbeam xrad xonly`. GLAD will treat the azimuthal direction correctly if `xonly` is specified.

The error in treating the near-cylindrical paths as strictly cylindrical is very small. Consider Fig. 17.10. A cross section of the beam is shown. If we constructed a beam with a perfectly uniform amplitude at the start, then when the beam moves to the outer cylindrical element, the outer part of the beam, marked A, will expand less than the inner part, marked B. Given that the wavefront is tilted by Δx across the beam, then the ratio of amplitude A to amplitude B, after propagation to the outer cylinder, is

$$\text{relative amplitude } \frac{A}{B} \approx 1 + \frac{\Delta x}{2}\left(\frac{1}{R_1} - \frac{1}{R_2}\right), \tag{17.2}$$

when we consider only the geometrical propagation. The effect would be twice this for intensity. This asymmetry will alter the diffraction propagation only slightly. When the beam returns to the inner cylinder, a geometrical representation would see the uniform amplitude restored. From this argument we see that the effect of beam tilt is small for modest angles from strictly radial propagation.

17.2.1 Decentration and Tilts

Small decentration of the elements, in cylindrical wave propagation, leads to an aberration which is constant axially and which varies azimuthally around the circumference. Small tips will introduce a phase error that varies axially and has the same azimuthal variation as decentration. Figure 17.11a illustrates tilt aberration of a radial beam and Fig. 17.11b illustrates decenter aberration of a radial beam.

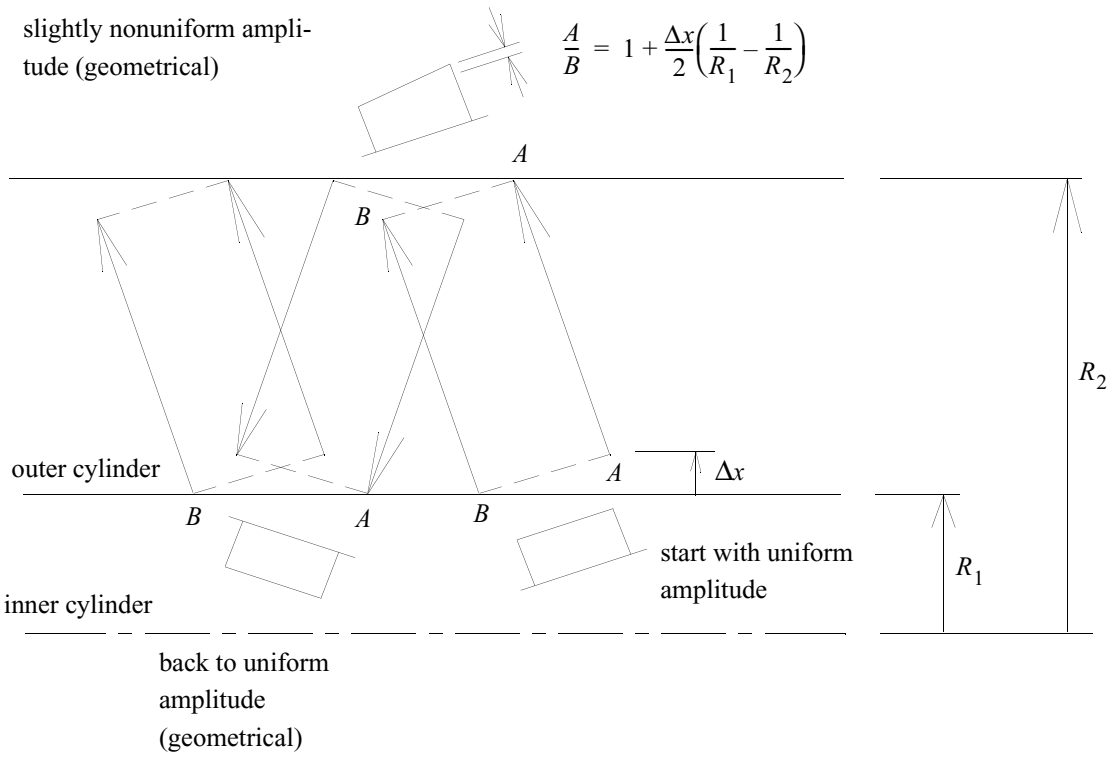


Fig. 17.10. A nearly cylindrical beam is shown reflecting between two cylindrical mirrors. The two edges of the beam are indicated by A and B. Edge A expands less than edge B in the upward path, leading to a slightly greater amplitude (the amplitude at the top has been renormalized for better legibility). This asymmetry will effect the diffraction propagation very slightly. Geometrically, the asymmetry is removed when the beam returns to the inner cylinder.

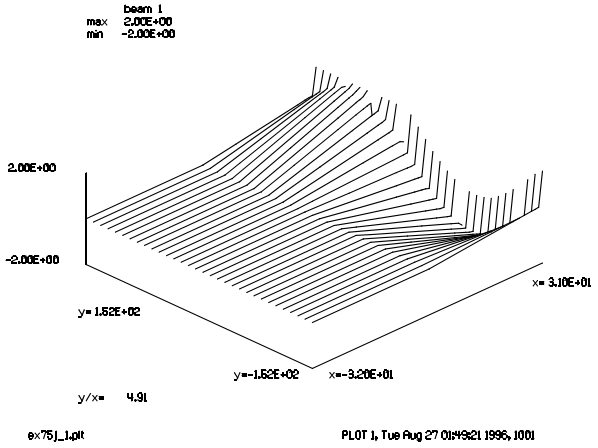


Fig. 17.11a. Phase plot of tilt aberration on a radial beam. The vertical direction corresponds to the azimuthal direction.

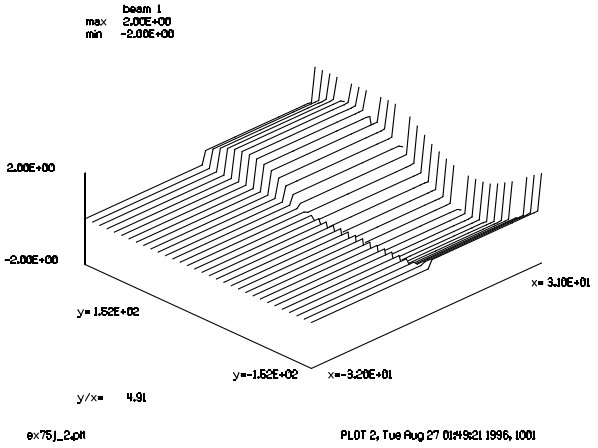


Fig. 17.11b. Phase plot of decenter aberration on a radial beam.

17.3 Polarization

When light is expanded into a radial beam by an axicon, the effect on polarization is to be rotated with the azimuth angle. If the light is reflected by a mirror in the radial path and returned to axial mode by a cone pointing in the same direction as the original cone, the polarization rotation is canceled and the polarization is well behaved. If the reconstructing cone is of opposite parity, the regenerated axial beam has its polarization state rotated at twice the azimuthal angle with respect to the initial beam.

Consider taking an axial beam, going through any combination of axicons, and returning the beam to axial form -- pointing in either direction. If the beam is not inverted by the optical path, then the polarization state will be well behaved. This is readily understood from basic image rotation properties.

From the previous discussion, it is easily seen that when viewing the axicon ray path in a y-z plane, the x-direction is always unchanged. If the y-direction is unchanged, i.e., the beam in not inverted by the axicon paths, then the system mirror matrix must take the form of

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \pm 1 \end{bmatrix}. \tag{17.3}$$

Either of these mirror systems has the a positive determinate for the (3,3) cofactor (see Table 11.1) and is insensitive to rotation about the z-axis. The polarization will not be rotated. Not that the rotation properties of refractive image inverting elements should not be considered in calculating the system calculation because refractive elements do not change the polarization with the image. Many axicon systems will have intermediate positions in which there is image rotation. If the diffraction distance in these regions is short, then little harm may be done. If however the beam goes to the far-field with a disrupted polarization state, the intensity pattern in the far-field will be greatly disturbed. If the polarization is rotating azimuthally, there will be zero intensity in the center of the far-field.

17.4 Annular Beams

An axicon acting on a radial beam will create an annular beam with a cylindrical axis. The special case when the radius of the cylindrical axis goes to zero the annular beam degenerates into an axial beam. If the beam has no significant optical divergence, then an afocal model may be applied. An annular beam which is essentially afocal may be treated as an axial beam for further propagation. The radial-to-annular transformation may be treated as a remapping of the radial intensity distribution and the optical axis redefined to be a simple line axis. GLAD uses an “afocal” operation mode to treat axicon operations as being afocal. In afocal operation GLAD does not calculate the transformation of the surrogate beam parameters. Also, shifts of the axicon along the axis are allowed. Fig. 17.12 illustrates the case of a left-shifted second cone, which produces a ring-like intensity pattern. A right shift of the second cone produces a peak in the intensity pattern of the regenerated axial beam. The image rotation properties are unchanged so the polarization rotation properties are not affected by the longitudinal shifts of the axicons. Fig. 17.13.a shows the intensity and Fig. 17.13b shows the polarization.

If the second axicon is right-shifted, as shown in Fig. 17.14, the intensity is disrupted as illustrated in Fig. 17.15.a with polarization behavior as shown in Fig. 17.15b.

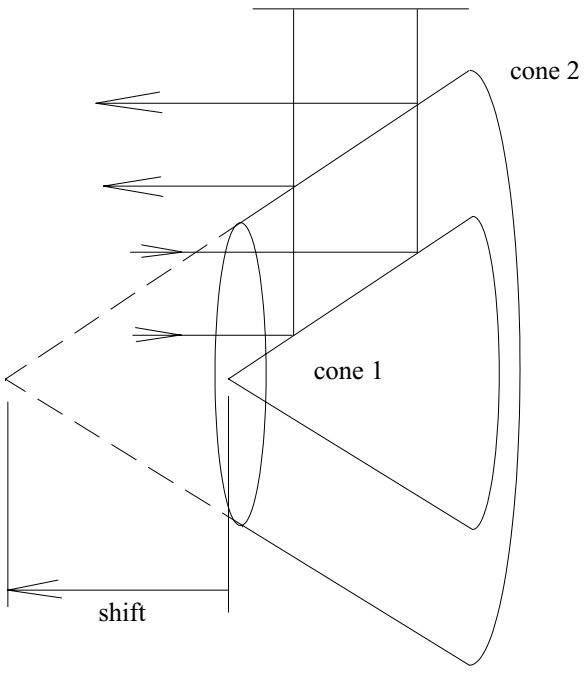


Fig. 17.12. The return axicon is shifted to the left causing a donut shaped beam to be formed.

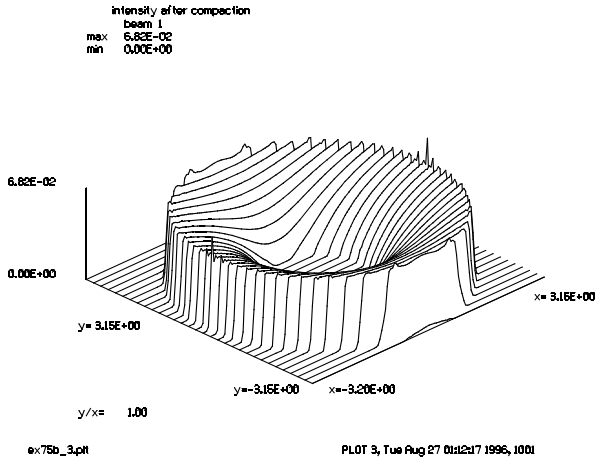


Fig. 17.13.a. Donut shaped beam formed by shifted axicon. Note small diffraction effects.

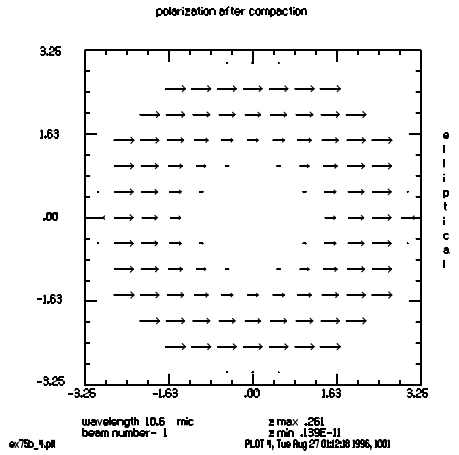


Fig. 17.13b. Polarization is returned to linear form.

If the return axicon is flipped over, as shown in Fig. 17.16, both the intensity and polarization will be disrupted. Axicons are designated as left or right based on which direction the apex points. The first axicon is left-shifted and the second is right-shifted so that the returning radial beam properly overlaps the second axicon. After the second axicon the optical axis is a cylinder, establishing that the beam is of annular form. The intensity is shown in Fig. 17.17a. The beam is inverted by the optical system, causing the polarization state to be rotated, as shown in Fig. 17.17b. The rotation of polarization causes the far-field pattern to have zero intensity on-axis as shown in Fig. 17.17c.

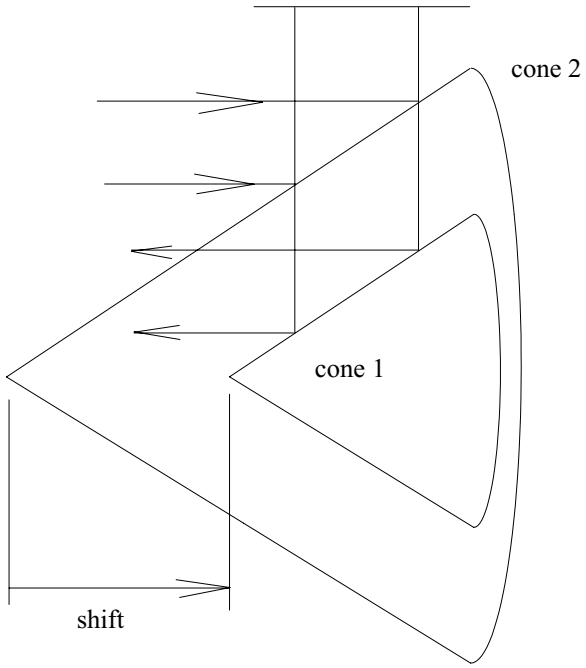


Fig. 17.14. In this example, a left axicon expands the beam into radial mode and the beam is returned to axial form by a right cone, leading to nonuniform intensity and disrupted polarization.

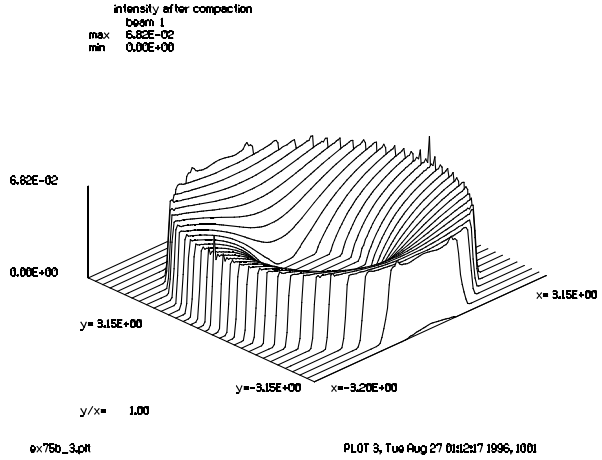


Fig. 17.15.a. Intensity after second left axicon. The shift to the right has resulted in nonuniform intensity.

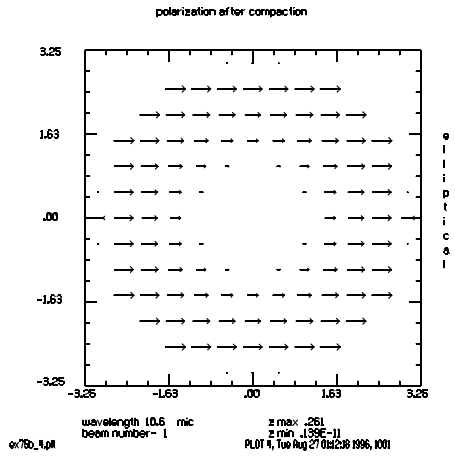


Fig. 17.15b. The polarization is well behaved unlike the case of left and right axicon mirrors.

Optical beams with significant optical power may be propagated through axicon system, as illustrated in Fig. 17.18. Radial mode presents no special difficulties, except that a second axicon will, in general, create an annular beam with a toroidal wavefront. GLAD uses a “focal” operating mode to be defined so that a radial beam with optical power may be properly returned to axial form. In focal operation GLAD will calculate the axicon axial position and left or right direction to properly return radial mode to axial mode with the `axicon/axial` command.

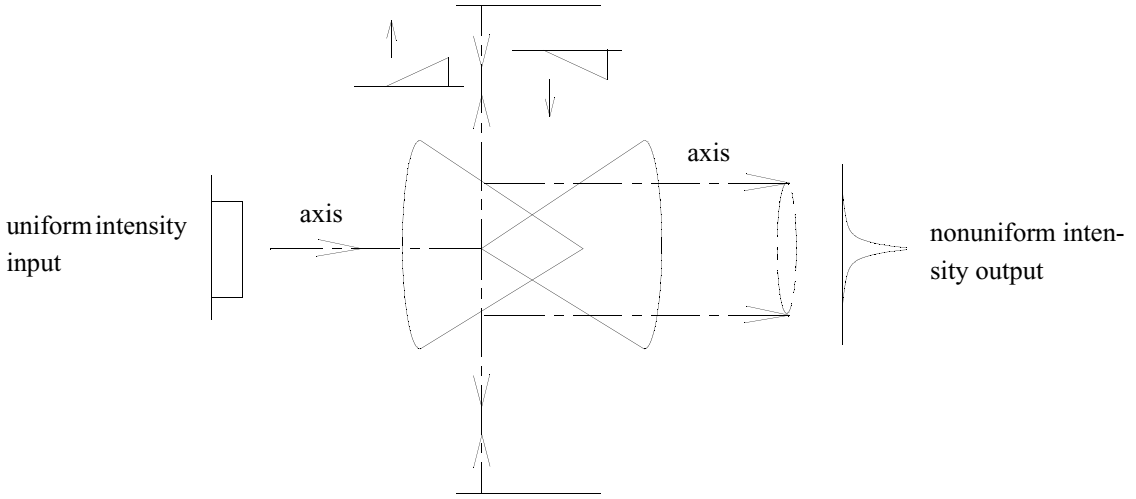


Fig. 17.16. In this example, a left axicon expands the beam into radial mode and the beam is returned to axial form by a right cone, leading to nonuniform intensity and disrupted polarization.

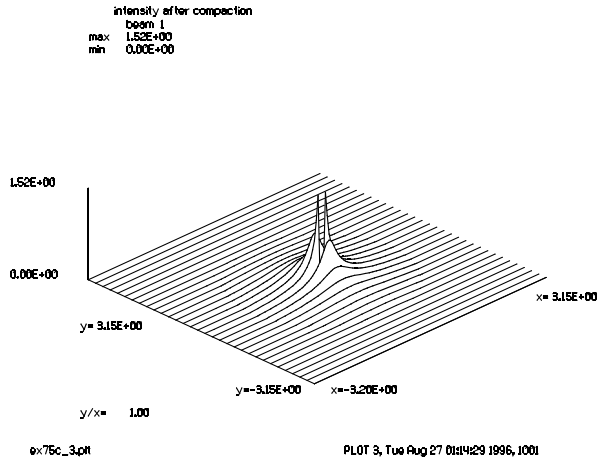


Fig. 17.17a. The intensity is strongly peaked in the center because of the use of the right cone rather than the same left cone.

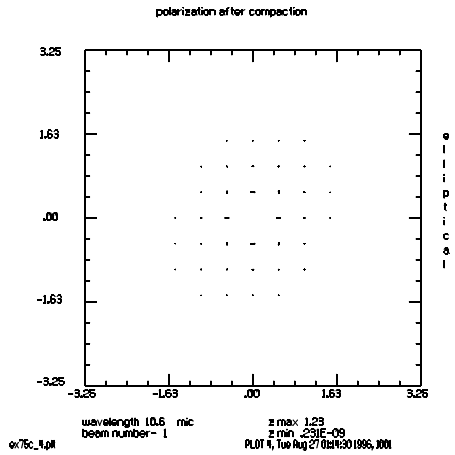


Fig. 17.17b. The polarization is strongly disrupted by the use of a right axicon in the return path.

In focal operation, the command `axicon/annular` will generate an annular beam with a toroidal wavefront. Consider Fig. 17.19 with a general axial-to-radial-to-annular conversion. GLAD keeps track of the reference position and the surrogate gaussian properties (in FOCAL operation) through the system. In radial mode the reference position is measured with respect to the original optical axis.

The early work in annular resonators used a fast-Hankel transform propagator. This propagator is useful only for essentially afocal. It allows propagation of axial beams which are radially symmetric or nearly radially symmetric. In the mid 1970's diffraction calculations were so expensive that going to a radially symmetric propagator made sense. In the 1990's large arrays could be propagated inexpensively and the need for a special radial propagator became less pressing. The Fast-Hankel transform is not as convenient as the FFT. The Fast-Hankel transform suffers from a singularity at the center, tends to “sag” in the outer

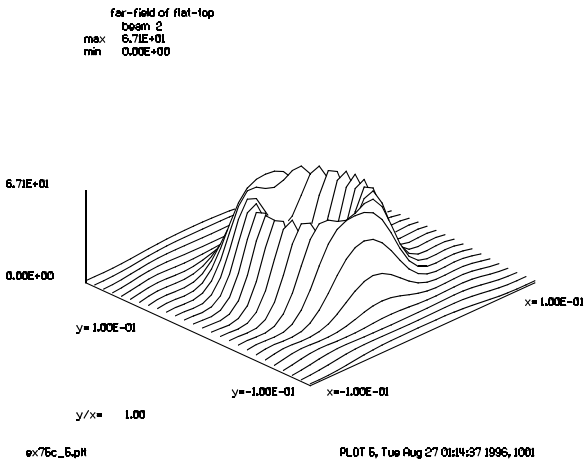


Fig. 17.17c. Expanded view of far-field image showing ring image due to polarization aberrations.

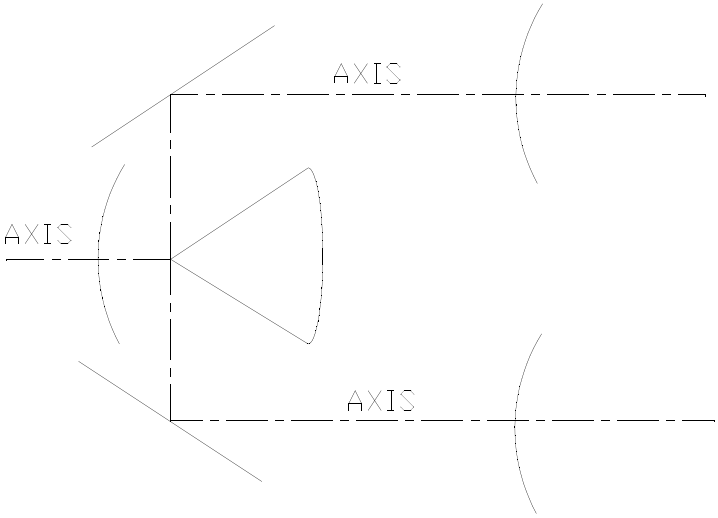


Fig. 17.18. A noncollimated beam may be propagated correctly in FOCAL operation. The divergence of the beam is propagated through the system and leads to a toroidal wavefront in annular mode.

regions, has nonuniform sampling, and is significantly slower than an equivalent size one-dimensional array. The fast-Hankel transform was incapable of treating strong toroidal wavefronts. A limited capability to represent azimuthal variation exists by maintaining and propagating several azimuthal slices. If more than few azimuthal slices are used, it becomes more efficient to use a standard square array.

The fast-Hankel transform was intended to model an annular laser with a chemical medium in the center. An annular beam propagating along a path with a large center obscuration is not properly represented by the Hankel transform. One should perform a split-step calculation with interposed steps of Hankel propagation and obscuration. In many ways the fast-Hankel transform solved the wrong problem. Because of the central obscuration, the intensity structure in the center of the beam, which the Hankel transform will predict, will not be allowed to form. What was needed was not a radially symmetric propagator but a torioidal propagator.

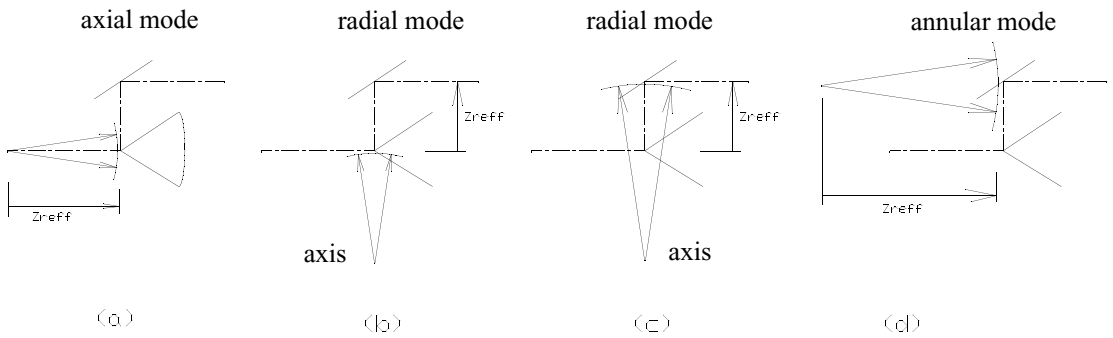


Fig. 17.19. In focal operation the reference position is measured radially with respect to the optical axis in radial mode. Fig. (a) shows the axial mode with Zreff = 0 to the left. In Fig. (b) Zreff is measured with respect to the optical axis. The center of curvature is shifted negatively. In Fig. (c) the beam has expanded to the next axicon mirror. In Fig. (d), Zreff is measured along the optical axis. The position Zreff = 0 has been displaced by the radius of the annular axis with respect to the optical axis.

The radial wave propagator is much closer to a toroidal wave propagation than is the Hankel transform. The toroidal wave, by stationary phase arguments, has diffraction effects primarily determined by the local complex amplitude. The radial propagator fits this geometry very well. The phase curvature of the toroidal wavefront prevents distant points from contributing to the local diffraction effects. The toroidal wavefront differs from the radial wavefront in that the contribution of the outer radial zones is greater than the inner radial zones, as shown in Fig. 17.20. This may be largely accounted for in diffraction propagation by area weighting the intensity before and after taking a diffraction step. This particularly important if the beam is flipped over by the toroidal equivalent of a spatial filter where the light goes through a ring focus, as demonstrated in Ex75i and illustrated in Fig. 17.21.

17.5 Polarization in Reflaxicon and Waxicon Configurations

The configuration of inner and outer reflective axicons has important effects on the polarization. In the case of an ordinary collimated beam with linear horizontal polarization incident on a left axicon mirror, a radial beam is created with polarization as shown in Fig. 17.22. Two left (or two right) reflecting axicons make a reflaxicon, which has similar optical characteristics as a rhomb prism. The polarization states rotates clockwise π when moving around the radial mode starting in the +y-direction. A second left axicon leaves the polarization state unchanged as displayed in radial-azimuth format and linearly polarized when shown in conventional form. A left-right pair (or right-left) pair of reflecting axicons constitute a waxicon configuration (so named because of the “w” shape created). This configuration is illustrated in Fig. 17.23, the radial mode properties are the same as for Fig. 17.22. Reflection from a right axicon flips the polarization about the radial direction and causes the rotation to change from $+2\pi$ to -2π, as displayed in radial-azimuth format. In conventional rectangular display, the polarization rotates -4π.

17.6 Resonators with Axicon Elements

Axicon elements have been suggested to take advantage of certain laser media which favors an annular gain regions. Many of the devices that were proposed in the 1970's were done without proper consideration of the properties of the optical axis and the image rotation. It is relatively easy to distinguish well behaved

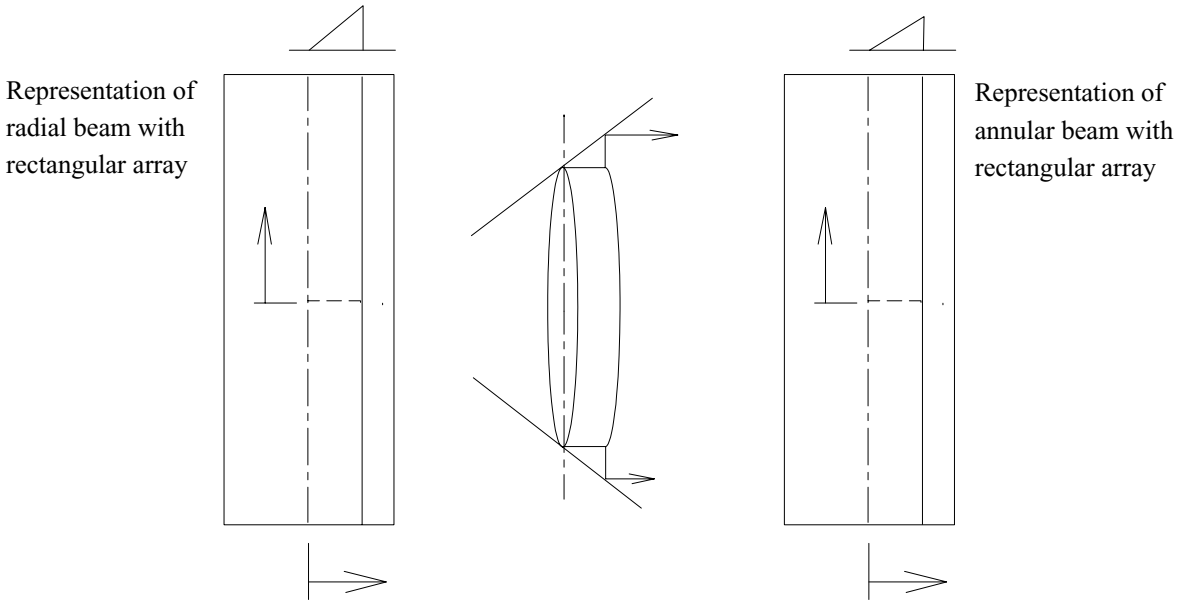


Fig. 17.20. Both radial and annular beams are represented in radial-azimuthal form with the azimuthal direction corresponding to the y-direction and filling the rectangular array with no guard band.

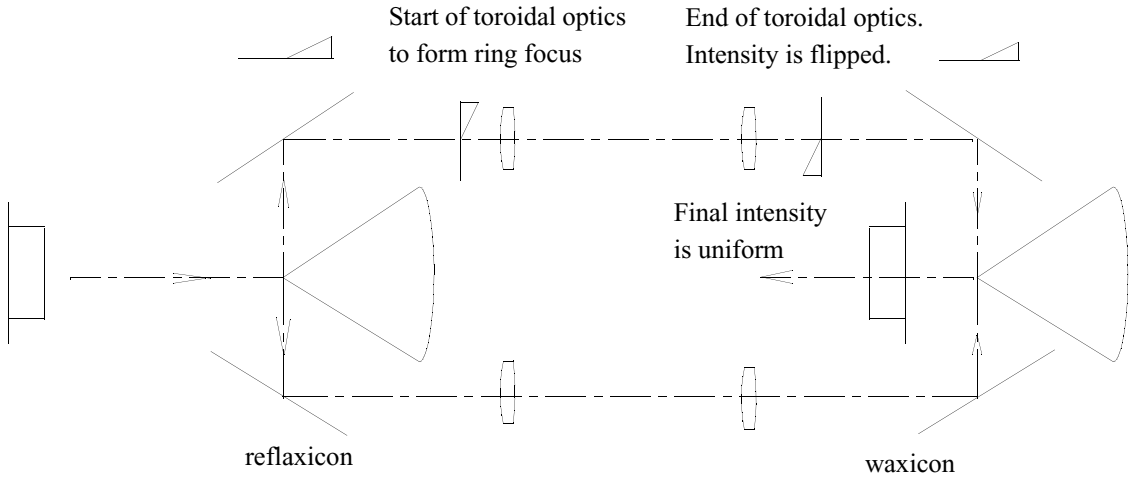


Fig. 17.21. Reflaxicon-waxicon with intermediate ring focus. The refractive optics in the annular path are toroidal lenses which flip the intensity about the cylindrical axis but do not change the polarization.

from poorly behaved devices simply on the basis of first order properties and the additional consideration of the need for limiting apertures to establish good mode selectivity. In a round trip through the device, we want the optical axis to return to axial form and also that there be no beam rotation, as viewed in the y-z plane. It should be noted that apertures in radial or annular mode, do not limit the azimuthal modes and, therefore, the azimuthal modes will be degenerate if there is no limiting in an axial mode. For good mode discrimination, it is essential to have a limiting aperture in an axial path.

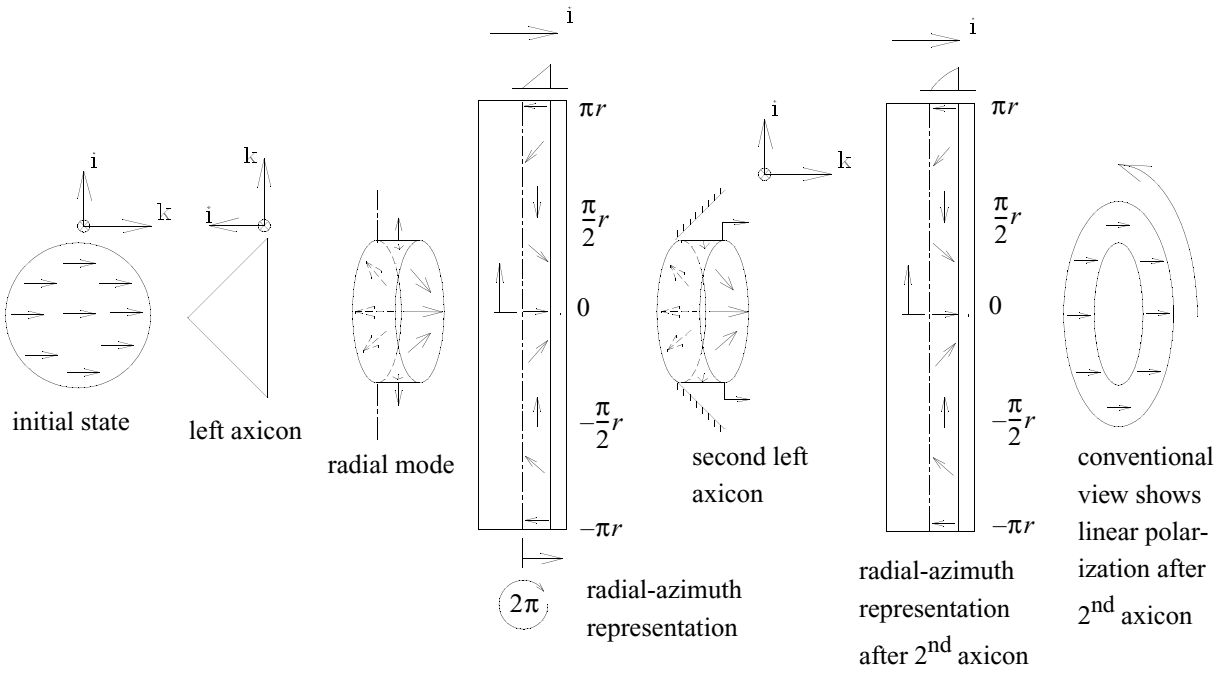


Fig. 17.22. A circular beam with horizontal linear polarization is reflected off a left axicon mirror, forming a radially expanding beam (radial mode). For radial-azimuth display, the beam is stretched into a rectangular array with the vertical direction spanning $-\pi r$ to $+\pi r$. The polarization state will be rotated counterclockwise for one cycle in the +y-direction. After reflection from a second left axicon, the polarization state will retain the same state of rotation and will be linear in conventional view.

There are numerous configurations resonators which have well behaved properties. Figs. 17.24 and 17.25 show two forms which have good properties. Two-axicon arrangements consisting of axicons of the same orientation (Fig. 17.24) are frequently called reflaxicons, but Iwaxicons (Fig. 17.25) if the axicons are of opposite orientation. Both configurations have good properties because the beam is not inverted (as observed in the y-z plane) in going through either system and there are limiting apertures in the axial legs. These systems will perform very similarly to the equivalent systems without axicon elements. The reflaxicon system has good polarization properties in the annular leg (Fig. 17.26). The polarization is aligned with the radius for all azimuthal positions, this gives the best propagation performance for toroidal beams.

For the waxicon system the polarization is rotated at twice the azimuthal angle (Fig. 17.27). Since the propagation effects in the annular leg are generally small, the disrupted polarization is not likely to cause significant effects. Most gain media is insensitive to the polarization state and the near field diffraction will be little affected by the variable polarization.

Normally a combination of a reflaxicon and an waxicon gives poor performance: nonuniform intensity and disrupted polarization. Figure 17.21 illustrates a reflaxicon-waxicon system with an intermediate ring focus that flips the beam. This corrects the image properties of the system and the intensity is uniform, as shown in Fig. 17.28. The polarization properties are not altered by the refractive system, so the polarization is disrupted in exactly the same form as without the ring focus (Fig. 17.29).

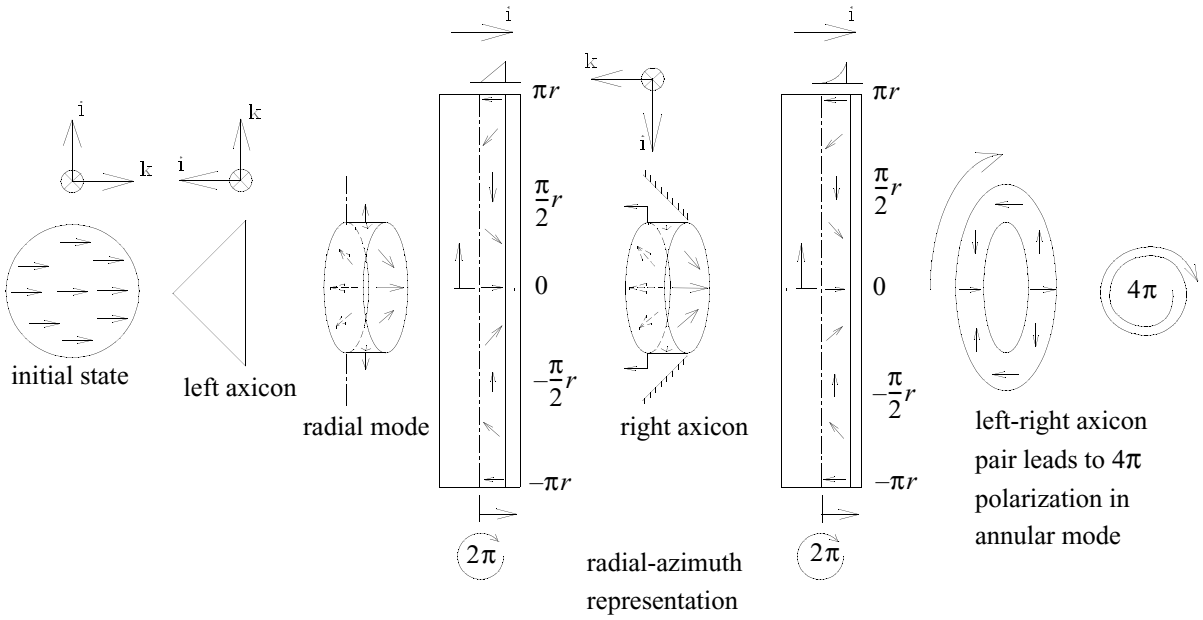


Fig. 17.23. Same as Fig. 17.22 except that the 2nd axicon is a right axicon and reverses the propagation direction. The true aperture shape now rotates counterclockwise leading to a 4π rotation when viewed in the conventional view for one cycle in the +y direction.

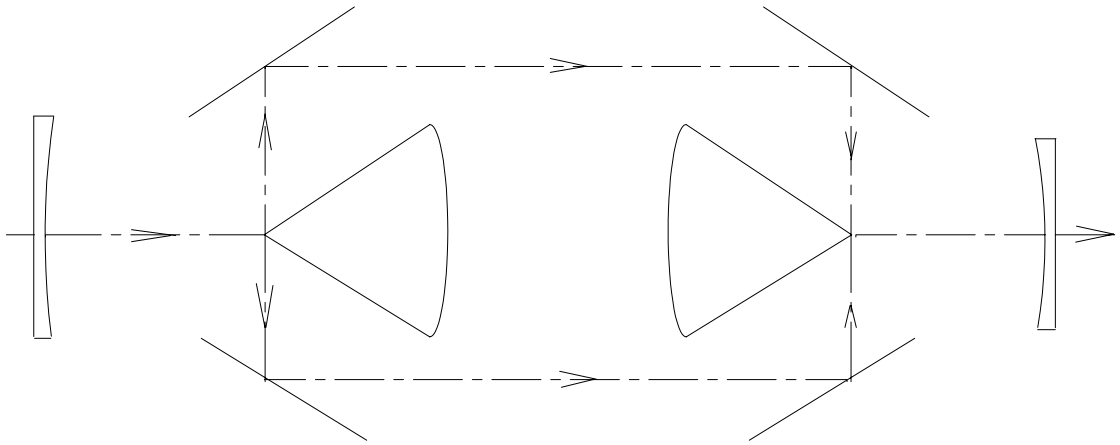


Fig. 17.24. Resonator configuration of two reflaxicons which generates an annular beam and after recompaction reforms as an axial beam.

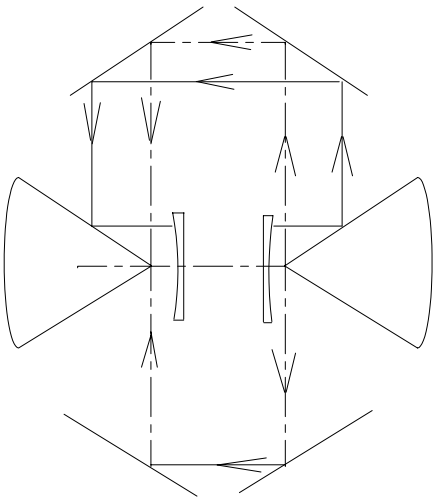


Fig. 17.25. Resonator configuration of two waxicons of opposite orientation. Note that the optical axis is on the outside of the annular region. A resonator could be made may putting concave or convex mirrors in the axial region.

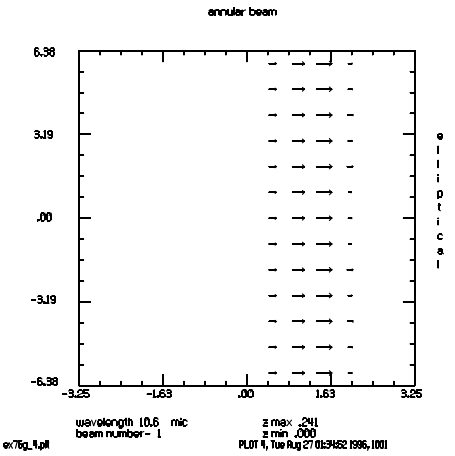


Fig. 17.26. Polarization is well behaved in annular path for reflaxicon of Fig. 17.22.

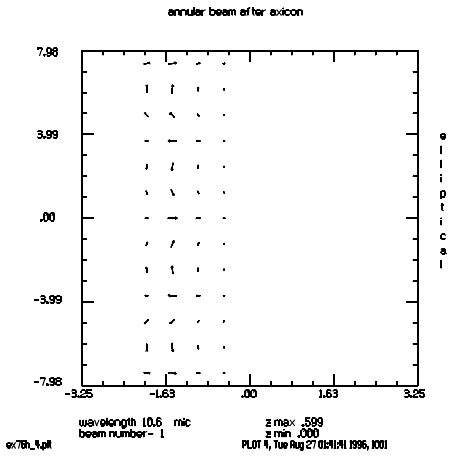


Fig. 17.27. Polarization in annular leg rotates at twice the azimuth angle for waxicon of Fig. 17.23.

17.7 More General Axicon Surfaces

Figure 17.30 shows a more general axicon configuration with powered axicons to introduce an internal focus in the radially converging leg. This subsystem flips the image so that the compacted leg will be properly reconstructed in both intensity and polarization state by the final right axicon. The optical axis is shown by the dashed line (center line style). The axis line is at the outer radius through the two waxicon pairs composing the front and back end-mirror assemblies.

This system requires more general mirrors than the simple 45º flat mirror profiles used earlier in this chapter. Axicon profiles with optical power and aspheric figure correction are implemented in GLAD by using a global mirror definition followed by an axicon irradiance remapping. The global mirror is used in

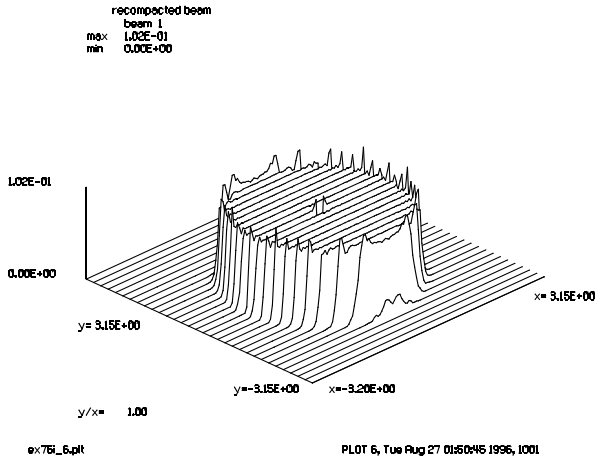


Fig. 17.28. Intensity in recompacted axial beam of system of Fig. 17.21 is the same as the starting form.

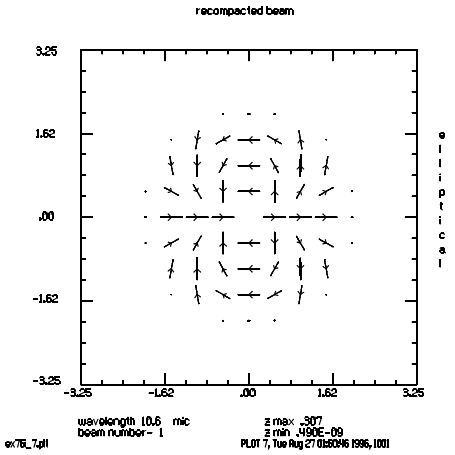


Fig. 17.29. The polarization in the recompacted axial beam after configuration of Fig. 17.21 is rotated at twice the azimuth angle.

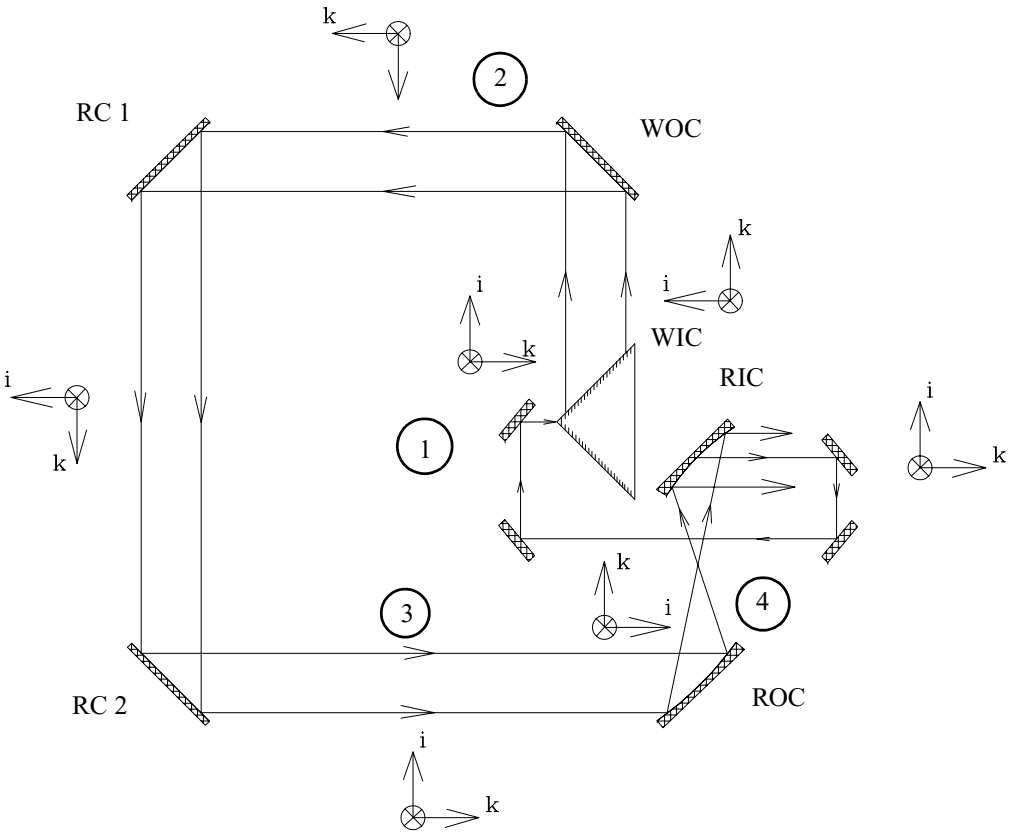


Fig. 17.30. Resonator configuration of two reflaxicons which generates an annular beam and after recompaction reforms as an axial beam.

the form `mirror/global/conic/radial` so that the aberration is computed only along a single radius in the x-direction and applied in all azimuthal directions equally. The axicon command is called with the `nomirror` parameter or parameter modifier to include only the axicon remapping. The mirrors should

be placed with the vertex at the axis line as shown in Fig. 17.30. The aberration polynomials should be defined in the vertex coordinate system of the mirror. Given an axial definition of the axicon surface of the form

$$z(x) = C_1 + C_2(x - x_0) + C_3(x - x_0)^2 + C_4(x - x_0)^3 + C_5(x - x_0)^4 + C_6(x - x_0)^5, \tag{17.4}$$

where x_0 is an offset of the polynomial description. We may equate this expression to a vertex coordinate system.

$$v = \frac{ct^2}{1 + [1 - c^2t^2(1 + \kappa)]^{1/2}} + a_4t^4 + a_4t^6 + \ldots . \tag{17.5}$$

Without loss of generality we may set the conic term to $\kappa = -1$ making the base surface parabolic with additional aspheric terms. Piston and tilt terms are added to better match the axis-based polynomial description:

$$v = B_0 + B_1t + \frac{1}{2}ct^2 + a_4t^4 + a_6t^6 . \tag{17.6}$$

The equation for the vertex line is, see Fig. 17.31:

$$z = C_1 + C_2(x - x_0) . \tag{17.7}$$

We can solve for the coefficients of Eq. (17.6) by the following steps (see Fig. 17.31):

1) For a number of values x, find $z(x)$ from Eq. 17.4.
2) Find $t = (x_t - Xdec)\cos\alpha + (z(x) - z_0)\sin\alpha$, where $\alpha = \tan^{-1}C_2$
3) Find v from Eq. (17.6) using t.
4) Find $z_t = t\sin(A) + v\cos(A) + z_0$
5) Minimize $(z(x) - z_t)^2$ for a range of x-points by solving for the coefficients of Eq. (17.6).

See Ex75k for an illustration of more general axicons.

References

1. A.E. Siegman, “Quasi fast Hankel transform,” Optics Lett. 1, 13–15 (March, 1977).
2. Shinan-Chur Sheng, Chap. 3, “Studies of Laser Resonators and Beam Propagation Using Fast Transfrom Methods,” PhD Dissertation, Dept. of Applied Physics, Stanford Univ., March 1980.
3. S-C. Sheng and A. E. Siegman, “Nonlinear optical calculations using fast transform methods: Second harmonic generation with depletion and diffraction,” Phys. Rev. A 21, 599-606 (February 1980).

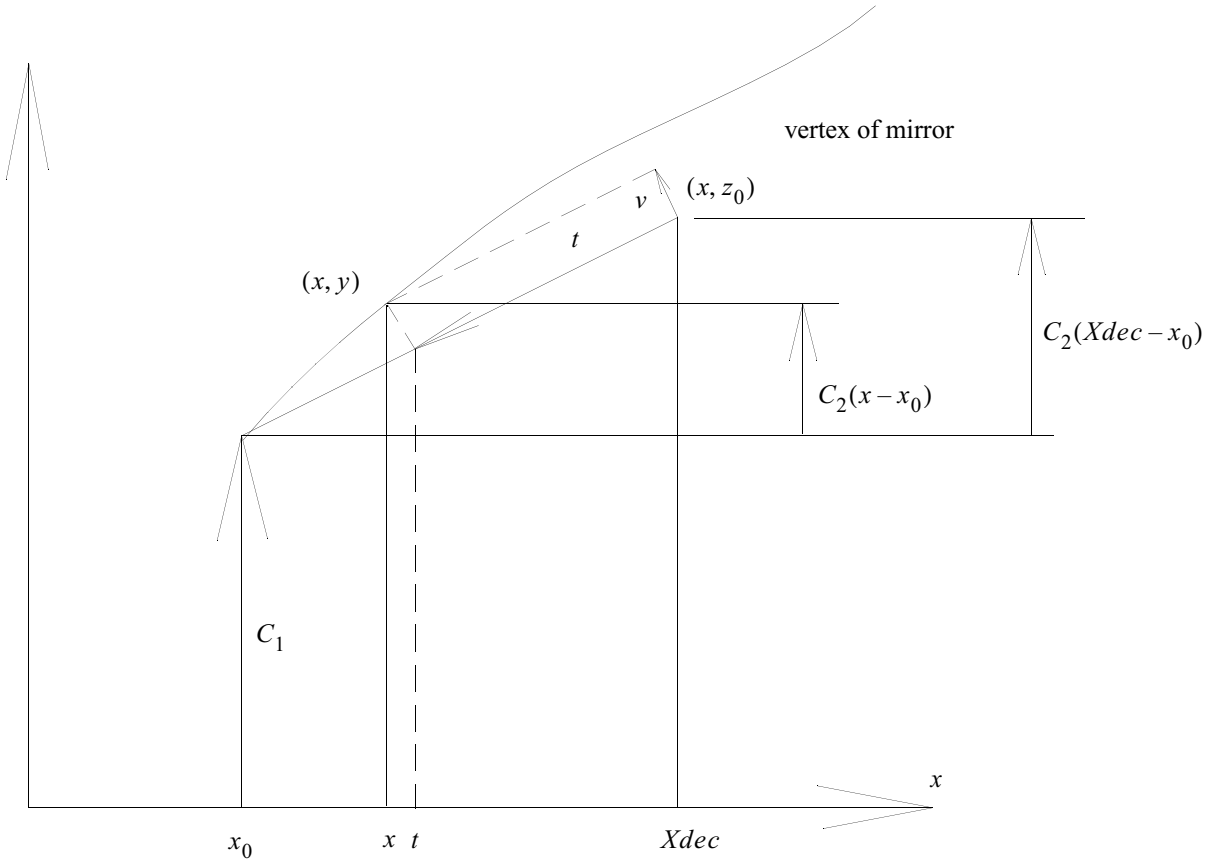


Fig. 17.31. An axicon surface may be defined by global coordinates from the point by global polynomials or from the vertex coordinate system at Xdec with coordinates.

4. A. V. Oppenheim, G. V. Frisk, and D. R. Martinez, “Computation of the Hankel transform using projections,” J. Acoust. Soc. Am. 68, 523-529 (1980).

5. D. Fink, “Polarization effects of axicons,” Appl. Opt. Vol. 18, No. 5, p581, March, 1979.

6. W. H. Southwell, “Multilayer coating design achieving a broadband 90 deg. phase shift,” Appl. Opt. Vol. 19, No. 16, p 2688, August 1980.

7. D. A. Holmes, P. D. Briggs, A. Prieto, A. Simonoff, K. C. Sun, and T. R. Waite, “Common pass decentered annular ring resonator (CPDARR),” SPIE Vol. 1224, p322 (1990).

8. Kenneth C. Sun, “Geometric analysis of reverse mode in decentered annular ring resonator (DARR),” SPIE Vol. 1224, p338 (1990).

9. C. M. Clayton and C. A. Huguley, “Experimental results of a common pass decentered annular ring resonator (CPDARR),” SPIE Vol. 1224, p359 (1990).

18. Partially Coherent Imaging

Partially coherent imaging has generally been investigated using geometrical optics. The speed of contemporary computers makes it quite reasonable to use a full diffraction analysis of a partially coherent imaging system, such as the one illustrated schematically in Fig. 18.1 and modeled in Example 83. The full diffraction analysis allows consideration of both near- and far-field diffraction. Using GLAD one may input essentially any type of optics and may observe the optical beam characteristics at any point in the system in a great variety of ways. As discussed below, no special modifications are needed to model partial coherence; the system is modeled as a standard coherent system, taking a single source point at a time, combined with a simple summation of the irradiance pattern due to the source point. The standard coherent propagation algorithms are thoroughly tested, so we may have confidence in the result.

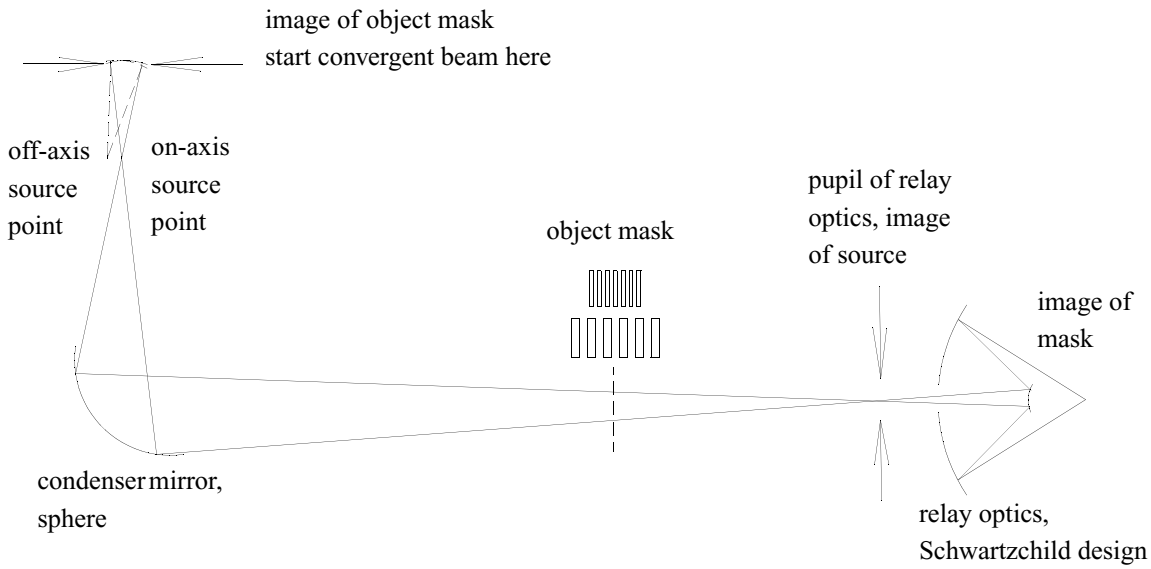


Fig. 18.1. The optical system consists of a condenser mirror, a tipped sphere, and a relay mirror of Schwartzchild design. The source is imaged into the pupil of the relay mirror. The beam is started at the conjugate of the object mask and is imaged by the condenser mirror to the object mask, where the transmission mask is located,. The beam is converging at the object mask and forms a point image at the pupil of the relay lens. A multi-bar pattern will form a central lobe and side lobes at the pupil of the relay lens. Different source points are simulated by tilting the initial beam where it is created at the conjugate of the object mask, as shown by the off-axis source point indicated as dotted lines.

While the calculation methods are simple, characterization of the errors in partially coherent imaging is somewhat difficult because, unlike purely coherent or purely incoherent imaging, we can not characterize the system by linear systems methodology such as the point response or frequency response. We must resort to examining the partial coherent images of selected objects of interest. The nonlinearity of partial coherent

imaging makes it difficult to be specific about accuracy and sampling requirements. We may, however, test the calculation method by evaluating it at the purely coherent and purely incoherent limits.

There are two methods which may be considered for performing partially coherent imaging: time or source area sampling. For time sampling, we consider that the source is a finite area of sufficient spectral bandwidth and is observed over some finite integration time. Using GLAD we can model finite spectral bandwidth for the finite source by using a moving temporal average of spatially delta-correlated random complex amplitude patterns. At each time sample, the complex amplitude distribution representing the current state of the source is propagated through the optical system to the image plane where an irradiance detector of finite bandwidth simulates the response of the optical recording medium. The finite bandwidth of the detector may be modeled by a moving, weighted temporal average of the irradiance distribution at the image plane. Numerical experiments with this type of direct temporal modeling using GLAD showed that for typical problems many hundreds of time samples were required to achieve satisfactory calculation of incoherent images.

A more efficient method, developed by H. H. Hopkins and used by D. S. Goodman is to consider that the source consists of incoherent points [1–2]. We find the irradiance image of each source point separately and then sum all of these irradiance images. Since we are performing a numerical calculation we must use a finite number of sample points for the source. Let the source points be represented by delta functions of the form $\delta(x_i, y_i)$, the optical system by an operator **H**, the complex amplitude at the image plane associated with each source point by $a(x', y', x_i, y_i)$, and the time averaged irradiance at the image plane by $I(x, y)$. x_i and y_i are the coordinates of the ith source point. The operator **H** includes all apertures, optical elements, diffraction, object mask, etc. which constitute the optical system. This operator treats the light as being strictly coherent and, therefore, uses the well-tested coherent propagation features of GLAD. For each ith source point, the complex amplitude distribution in the image plane is determined. The final image plane irradiance is found by

$$\frac{1}{N}\sum\left|\mathrm{H}[\delta(x_i, y_i)]\right|^2 = \frac{1}{N}\sum\left|a(x', y', x_i, y_i)\right|^2 \approx I(x, y)\,. \tag{18.1}$$

Implementation of this equation is very easy: the coherent propagation represented by **H** uses standard features and may be evaluated by well-known Fourier optics methods, the incoherent sum is simple to perform. About the only ways to make a mistake are to define the optical prescription incorrectly (this can occur with any method) or to have inadequate source sampling. The simplicity of the Hopkins' source sampling method leads to a simple procedure for validation. We assess the performance of a single source point to be sure the system model has been described correctly, i.e., that we have defined the optical prescription properly. We can also evaluate the sufficiency of the source sampling by calculating the image of a source which fills the pupil of the relay lens ($\sigma = 1$). Consider circular regions of the source defined by a normalized source radius variable r_s

$$I(x, y) \approx \frac{1}{N}\sum\left|a(x', y', x_i, y_i)\right|^2 \tag{18.2}$$

$$\approx \int_0^{\varepsilon}\left|a(r_s, x, y)\right|2\pi r_s dr_s + \int_{\varepsilon}^{\sigma}\left|a(r_s, x, y)\right|2\pi r_s dr_s + \int_{\sigma}^{1}\left|a(r_s, x, y)\right|2\pi r_s dr_s\,.$$

The radial region from 0 to an infinitesimal radius ε corresponds to coherent illumination. The region from 0 to some finite radius σ (incorporating the coherent illumination region) corresponds to partially coherent imaging. Integration out to $r_s = 1$ corresponds to full relay pupil filling and gives essentially incoherent imaging. Since full filling of the pupil necessarily incorporates the partial filling regions corresponding to partial coherence, we believe a satisfactory incoherent image (calculated in this way) implies satisfactory partial coherence.

Because each object mask results in somewhat different sampling requirements, it is difficult to derive a comprehensive rule that will hold for all mask and source geometries. For a specific mask and source geometry, it is quite easy to determine the adequacy of a selected sampling density simply by trying a greater number of sample points. If denser sampling does not change the results of the calculation, the original sampling density is adequate. If we see little peaks on square bar patterns ostensibly imaged incoherently, we know that the sampling density is not sufficient. To accurately assess the errors for a given mask, we can calculate the $\sigma = 1$ distribution and compare it with convolution of the object mask with the incoherent point spread function.

Hopkins calculated the modulation in the image plane of rectangular bar objects of various periods using several σ values [1]. Hopkins assumed the rectangular bar target was of infinite extent. This has been approximated by using thirteen bars in Example 83c. We selected the period of the bar pattern to be at 0.5 of the limit of the coherent transfer function. This is the same as Hopkins $\sigma = 0.5$ parameter. Figure 18.2 shows the thirteen-bar mask. Figure 18.3 shows the diffraction pattern in the pupil of the relay lens due to a single source point. The side bands represent the fundamental frequency of the mask. For Hopkins's bar pattern of infinite width, the principle lobe and side bands are of infinitesimal width. Figure 18.4 shows the modulation in a single cycle for σ = 0, 0.7, and 1.0 in good agreement with Hopkin's result using an infinite length rectangular pattern [1]. We used the source sampling pattern illustrated in Example 83 for these calculations. About 49 sample points over the source area gives a good approximation to full incoherence.

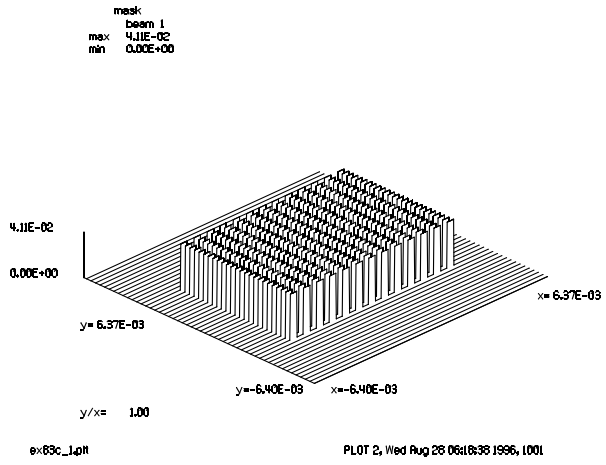


Fig. 18.2. Thirteen-bar target set to a period of 0.5 of the coherent diffraction limited spatial frequency.

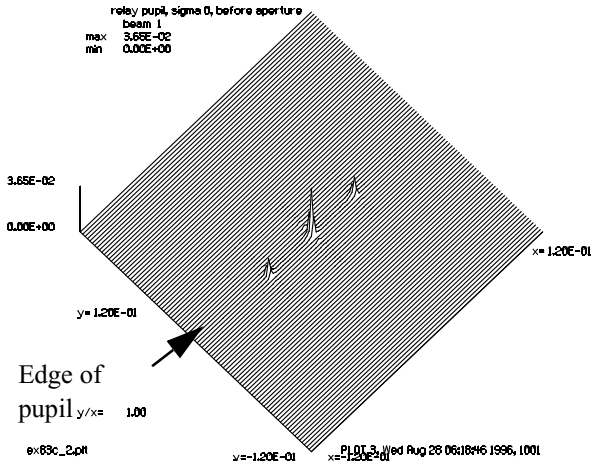


Fig. 18.3. Irradiance at relay pupil due to a single source point and the thirteen bar pattern with spatial frequency equal to 0.5 of the coherent diffraction limit.

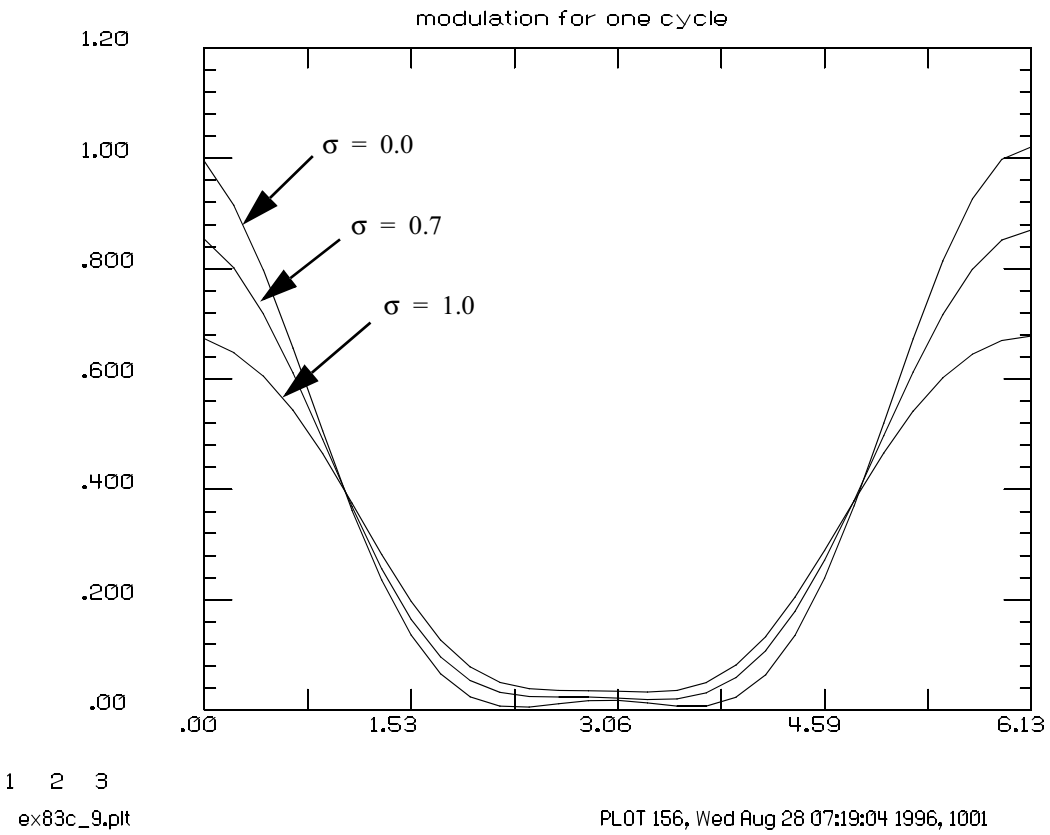


Fig. 18.4. Modulation at $\sigma = 0.0, 0.7, 1.0$ for 0.5 of the coherent diffraction limited frequency ($\omega = .5$), for a thirteen bar pattern. This compares very well with the $\omega = 0.5$ bar pattern of infinite length calculated by Hopkins [1].

References

1. H. H. Hopkins, “On the diffraction theory of optical images,” Proc. Roy. Soc. (London) A217, pp 408-432 (1953).

2. Douglas S. Goodman, “Lithographic image simulations,” Microelectronic 3 (1985) 355-362.

19. Thermal and Stress Effects in Component

GLAD can calculate the thermal response of refractive components including the effects of mounting materials and fluids in contact with the solid materials as well as the birefringent effects of mechanically and thermally induced stress.

19.1 Thermal Effects

The general expression for heat conduction in the presence of an internal heat source is

$$\rho c \frac{\partial T}{\partial t} = \nabla \cdot (\kappa \nabla T) + Q\,, \tag{19.1}$$

where ρ is the density, c is the heat capacity, T is the temperature, t is time, κ is the thermal conductivity, and Q is an internal heat source. In GLAD the medium is represented by a two-dimensional array of data or a collection of two-dimensional arrays to form a three-dimensional array. The rectangular arrangement of data establishes a natural rectangular node system for finite element solution.

Consider two small three-dimensional regions, which may be of different materials. See Fig. 19.1. Heat conduction for these two interacting regions obeys (approximately) the equations

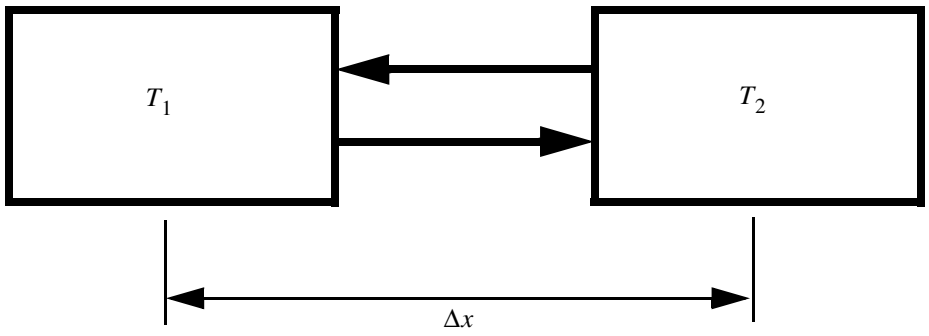


Fig. 19.1. Two interacting elemental regions in a rectangular node structure for finite element calculations.

$$\rho_1 c_1 \frac{\partial T_1}{\partial t} \approx \frac{\kappa}{\Delta x^2}(T_2 - T_1)\,, \tag{19.2}$$

$$\rho_2 c_2 \frac{\partial T_2}{\partial t} \approx \frac{\kappa}{\Delta x^2}(T_1 - T_2)\,,$$

where

$$\bar{\kappa} = \frac{2}{\frac{1}{\kappa_1} + \frac{1}{\kappa_2}} \tag{19.3}$$

is the mean thermal conductivity. By conservation of energy

$$T_1 \rho_1 c_1 + T_2 \rho_2 c_2 = E, \tag{19.4}$$

where E is the total energy. Conservation of energy allows us to separate the coupled differential equations:

$$\frac{\partial T_1}{\partial t} = \frac{\bar{\kappa}}{\Delta x^2}\left[\frac{E}{\rho_1 c_1 \rho_2 c_2} - T_1\left(\frac{1}{\rho_2 c_2} + \frac{1}{\rho_1 c_1}\right)\right], \tag{19.5}$$

$$\frac{\partial T_2}{\partial t} = \frac{\bar{\kappa}}{\Delta x^2}\left[\frac{E}{\rho_1 c_1 \rho_2 c_2} - T_2\left(\frac{1}{\rho_2 c_2} + \frac{1}{\rho_1 c_1}\right)\right].$$

Equation 19.5 may be written in the form $\partial T/\partial t = a + bT$ with the solution $T(t) = [T(0) + a/b]\exp(bt) - a/b$. We have the solutions,

$$T_1(t) = \frac{\rho_2 c_2}{\rho_1 c_1 + \rho_2 c_2}[T_1(0) - T_2(0)]e^{-\frac{\bar{\kappa}}{\Delta x^2}\frac{\rho_1 c_1 + \rho_2 c_2}{\rho_1 c_1 \rho_2 c_2}t} + \frac{T_2(0)\rho_2 c_2 + T_1(0)\rho_1 c_1}{\rho_1 c_1 + \rho_2 c_2}, \tag{19.6}$$

$$T_2(t) = \frac{\rho_1 c_1}{\rho_1 c_1 + \rho_2 c_2}[T_2(0) - T_1(0)]e^{-\frac{\bar{\kappa}}{\Delta x^2}\frac{\rho_1 c_1 + \rho_2 c_2}{\rho_1 c_1 \rho_2 c_2}t} + \frac{T_2(0)\rho_2 c_2 + T_1(0)\rho_1 c_1}{\rho_1 c_1 + \rho_2 c_2}.$$

A set of finite difference equations, which conserve energy, are

$$T_1(\Delta t) - T_1(0) = \Delta T_1 = \frac{\rho_2 c_2}{\rho_1 c_1 + \rho_2 c_2}[T_2(0) - T_1(0)]\left(1 - e^{-\frac{\bar{\kappa}}{\Delta x^2}\frac{\rho_1 c_1 + \rho_2 c_2}{\rho_1 c_1 \rho_2 c_2}\Delta t}\right), \tag{19.7}$$

$$\Delta T_2 = -\Delta T_1 \frac{\rho_1 c_1}{\rho_2 c_2}. \tag{19.8}$$

For rotationally symmetrical problems, consider two thin annular regions, which may be of different materials. Heat conduction obeys the equation

$$\rho c\frac{\partial T}{\partial t} = \bar{\kappa}\left(\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}\right)T. \tag{19.9}$$

The derivatives may be expressed in term of the previous point T_0 (left), current point T_1 (center), and T_2 (right). We have the approximate expressions

$$\frac{\partial T}{\partial t}\approx\frac{T_2-T_0}{2\Delta r} = \frac{T_2-T_1}{2\Delta r}-\frac{T_0-T_1}{2\Delta r}, \tag{19.10}$$

$$\frac{\partial^2 T}{\partial r^2}\approx\frac{T_2+T_0-2T_1}{\Delta r^2} = \frac{T_2-T_1}{2\Delta r^2}+\frac{T_0-T_1}{2\Delta r^2}. \tag{19.11}$$

Now considering only the interaction between T_1 and T_2 (interaction between T_0 and T_1 has already been considered when T_0 was the current point), we have

$$\rho_1 c_1\frac{\partial T_1}{\partial t}\approx\frac{\bar{\kappa}}{\Delta r^2}\left(1+\frac{\Delta r}{2r}\right)(T_2-T_1), \tag{19.12}$$

$$\rho_2 c_2\frac{\partial T_2}{\partial t}\approx-\frac{\bar{\kappa}}{\Delta r^2}\left(1-\frac{\Delta r}{2r}\right)(T_2-T_1).$$

In finite difference form, Eq. (19.12) becomes

$$\Delta T_1\approx\frac{\bar{\kappa}}{\rho_1 c_1}\frac{1}{\Delta r^2}\left(1+\frac{\Delta r}{2r}\right)(T_2-T_1)\Delta t, \tag{19.13}$$

$$\Delta T_2\approx-\frac{\bar{\kappa}}{\rho_2 c_2}\frac{1}{\Delta r^2}\left(1-\frac{\Delta r}{2r}\right)(T_2-T_1)\Delta t.$$

Equation (19.13) does not strictly conserve energy. Conservation of energy requires,

$$\Delta E = \rho_2 c_2 r_2\Delta T_2+\rho_1 c_1 r_1\Delta T_1 = 0, \tag{19.14}$$

but Eq. (19.13) yields a finite energy error of

$$\Delta E = \Delta t\frac{\bar{\kappa}}{\Delta r^2}(T_2-T_1)\left[-(r_1+\Delta r)\left(1-\frac{\Delta r}{2r_1}\right)+r_1\left(1+\frac{\Delta r}{2r_1}\right)\right] = \Delta t\bar{\kappa}\frac{(T_2-T_1)}{r_1}. \tag{19.15}$$

This problem is corrected by making the approximations

$$1+\frac{\Delta r}{2r}\approx\sqrt{\frac{r_2}{r_1}}\text{ and }1-\frac{\Delta r}{2r}\approx\sqrt{\frac{r_1}{r_2}}, \tag{19.16}$$

$$\rho_1 c_1\frac{\partial T_1}{\partial t}\approx\frac{\bar{\kappa}}{\Delta r^2}(T_2-T_1)+\frac{\bar{\kappa}(T_2-T_1)}{2r\Delta r}\approx\frac{\bar{\kappa}}{\Delta r^2}\sqrt{\frac{r_2}{r_1}}(T_2-T_1), \tag{19.17}$$

$$\rho_2 c_2\frac{\partial T_2}{\partial t}\approx\frac{\bar{\kappa}}{\Delta r^2}(T_1-T_2)+\frac{\bar{\kappa}(T_1-T_2)}{2r\Delta r}\approx-\frac{\bar{\kappa}}{\Delta r^2}\sqrt{\frac{r_1}{r_2}}(T_2-T_1). \tag{19.18}$$

Conservation of energy, Eq. (19.14), allows us to separate the coupled differential equations:

$$\frac{\partial T_1}{\partial t}=\frac{\bar{\kappa}}{\Delta r^2}\sqrt{\frac{r_2}{r_1}}\left[\frac{E}{\rho_1 c_1\rho_2 c_2 r_1}-T_1\left(\frac{r_1}{r_2}\frac{1}{\rho_2 c_2}+\frac{1}{\rho_1 c_1}\right)\right], \tag{19.19}$$

$$\frac{\partial T_2}{\partial t}=\frac{\bar{\kappa}}{\Delta r^2}\sqrt{\frac{r_1}{r_2}}\left[\frac{E}{\rho_1 c_1\rho_2 c_2 r_2}-T_2\left(\frac{r_2}{r_1}\frac{1}{\rho_1 c_1}+\frac{1}{\rho_2 c_2}\right)\right],$$

$$T_1(t)=\frac{\rho_2 c_2 r_2}{\rho_1 c_1 r_1+\rho_2 c_2 r_2}[T_1(0)-T_2(0)]e^{-\frac{\bar{\kappa}}{\Delta r^2}\left(\sqrt{\frac{r_2}{r_1}}\frac{1}{\rho_1 c_1}+\sqrt{\frac{r_1}{r_2}}\frac{1}{\rho_2 c_2}\right)t}+\frac{T_2(0)\rho_2 c_2 r_2+T_1(0)\rho_1 c_1 r_2}{\rho_1 c_1 r_1+\rho_2 c_2 r_2}, \tag{19.20}$$

$$T_2(t)=\frac{\rho_1 c_1 r_1}{\rho_1 c_1 r_1+\rho_2 c_2 r_2}[T_2(0)-T_1(0)]e^{-\frac{\bar{\kappa}}{\Delta r^2}\left(\sqrt{\frac{r_2}{r_1}}\frac{1}{\rho_1 c_1}+\sqrt{\frac{r_1}{r_2}}\frac{1}{\rho_2 c_2}\right)t}+\frac{T_2(0)\rho_2 c_2 r_2+T_1(0)\rho_1 c_1 r_2}{\rho_1 c_1 r_1+\rho_2 c_2 r_2}.$$

A set of finite difference equations, which conserve energy, are

$$T_1(\Delta t)-T_1(0)=\Delta T_1=\frac{\rho_2 c_2 r_2}{\rho_1 c_1 r_1+\rho_2 c_2 r_2}[T_2(0)-T_1(0)]\left[1-e^{-\frac{\bar{\kappa}}{\Delta r^2}\left(\sqrt{\frac{r_2}{r_1}}\frac{1}{\rho_1 c_1}+\sqrt{\frac{r_1}{r_2}}\frac{1}{\rho_2 c_2}\right)\Delta t}\right], \tag{19.21}$$

$$\Delta T_2=-\Delta T_1\frac{\rho_1 c_1 r_1}{\rho_2 c_2 r_2}. \tag{19.22}$$

The time constant for a given material with a particular spatial sampling is

$$\tau=\frac{\rho c\Delta x^2}{\bar{\kappa}}. \tag{19.23}$$

For a given choice of spatial sampling, the materials will differ by the ratio $\rho c/\kappa$, with thermal influence function $\exp(-\rho c r^2/\kappa t)$. When different materials are in contact, the temporal sampling must be sufficiently short to resolve the shortest time constant.

In GLAD, fluids are assumed to maintain their temperature independently of heat flow across the boundary. As such, a fluid is characterized by its temperature and convection coefficient. Most elements are in contact with air. Other fluids may be forced flow air, water, forced flow water, etc. The temperature change due to convection obeys the equation,

$$\frac{\partial T}{\partial t} = \frac{h}{\rho c \Delta x}(T_\infty - T), \tag{19.24}$$

where Δx is the length of the incremental region (perpendicular to the convection surface)

$$T(t) = (T(0) - T_\infty)e^{-\frac{h}{\rho c \Delta x}t} + T_\infty, \tag{19.25}$$

$$\Delta T = (T_\infty - T(0))\left(1 - e^{-\frac{h}{\rho c \Delta x}\Delta t}\right). \tag{19.26}$$

The time constant for convection is

$$\tau = \frac{\rho c \Delta x^2}{h}, \tag{19.27}$$

similar to Eq. 19.23.

In GLAD, a number of materials may be combined to represent glass (or similar transparent material), opaque material for mounting materials, and fluid such as air or water acting as a heat sink. Fig. 19.2 indicates schematically material representation.

19.2 Thermo-Optical Effects

Nonuniform temperature distributions induce phase effects in windows and lenses due to local changes in the index of refraction and expansion coefficient. To second order, the index of refraction can be represented by

$$N(T) = N(0) + \frac{dN}{dt}(T - T_0) + \frac{1}{2}\frac{d^2N}{dt^2}(T - T_0)^2 = N(0) + \Delta N, \tag{19.28}$$

where $N(T)$ is the index of refraction and T_0 is the reference wavelength. The change in thickness of the element is determined to second order by,

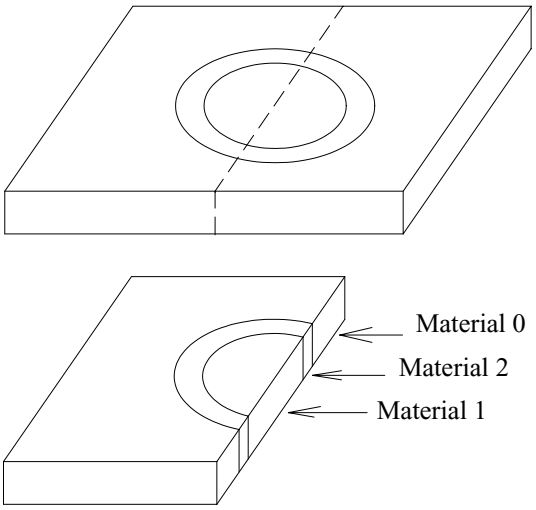


Fig. 19.2. The medium for thermal analysis is divided into sections consisting of rectangular arrays. A single section is shown above. In this case, the center region is the optical material, Material 2 is a mounting material such as aluminum, the outer material may be air. A thick element may be represented by several such sections. Solid materials may have internal heat sources. The laser beam passing through the refractive material can also inject heat into the optical material due to absorption. Thermal conduction redistributes the heat transversely within a section and axially between adjacent sections. Thermal convection causes heat to be lost into the air. The mesh is build using the standard GLAD array architecture.

$$\frac{\Delta L}{L} = \frac{1}{L}\frac{dL}{dT}(T - T_0) + \frac{1}{2}\frac{1}{L}\frac{d^2L}{dT^2}(T - T_0)^2 . \tag{19.29}$$

The term dL/LdT is the thermal coefficient of expansion. The optical phase due to temperature changes in a window is

$$\theta = \frac{2\pi}{\lambda}\Big[(N(0) - 1)\Delta L + \Delta N L\Big]. \tag{19.30}$$

The optical power of a surface is

$$\phi = \phi_0 + \Delta\phi(T) = \phi_0\left(1 + \frac{\Delta\phi(T)}{\phi_0}\right), \text{ where } \phi_0 = \frac{N(0) - 1}{R}. \tag{19.31}$$

Including the differential effects of index and radius of curvature from Eqs. (19.28) and (19.29) respectively we have the relative change in optical power of a surface,

$$\frac{\Delta\phi(T)}{\phi_0} = \left(\frac{dN}{dT}\frac{1}{N(0)-1} - \frac{1}{L}\frac{dT}{dL}\right)(\langle T\rangle - T(0)), \tag{19.32}$$

where $\langle T\rangle$ is the irradiance-averaged temperature.

The optical material may have bulk absorption such that the irradiance is decreased according to the equation

$$I(L) = I(0)e^{-\alpha L}. \tag{19.33}$$

The change in temperature, averaged over the axial length L of the section, is

$$\Delta T = I(0)(1-e^{-\alpha L})\frac{\Delta t}{\rho c L}, \tag{19.34}$$

where Δt is the time of interaction.

19.3 Thermally Induced Birefringent Stress

Temperature changes may induce stress in optical elements. We shall assume that the component is in a properly designed kinematic mount so that the mount does not impose stress on the part. For example a window may be held supported in a compliant mount that allows for thermal expansion. A mirror may be supported with the reflective face registered to three hard points which constrain the face with respect to tip and tilt but do not resist expansion or contraction. Consequently thermally induced stresses are generated within the optical element and from the boundaries of the element. This allows for a simplification of the calculation in that thermally induced strains do not cause shear stresses.

A section of solid material will distort because of imposed stress (outward pressure, force per unit area), indicated in Fig. 19.3. Stresses induces strain (displacement per unit length, dimensionless) according to the tensor equation [1]:

$$\begin{bmatrix}\varepsilon_x\\ \varepsilon_x\\ \varepsilon_z\\ \gamma_{xy}\\ \gamma_{xz}\\ \gamma_{yz}\end{bmatrix} = \frac{1}{E}\begin{bmatrix}1 & -\mu & -\mu & 0 & 0 & 0\\ -\mu & 1 & -\mu & 0 & 0 & 0\\ -\mu & -\mu & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 2(1+\mu) & 0 & 0\\ 0 & 0 & 0 & 0 & 2(1+\mu) & 0\\ 0 & 0 & 0 & 0 & 0 & 2(1+\mu)\end{bmatrix}\begin{bmatrix}\sigma_x\\ \sigma_y\\ \sigma_z\\ \tau_{xy}\\ \tau xz\\ \tau_{yz}\end{bmatrix}. \tag{19.35}$$

where $\boldsymbol{\varepsilon}$ is the strain vector (with the γ's being the shear strain components), $\boldsymbol{\sigma}$ is the imposed stress vector (with the τ's being the shear stress components), E is Young's Modulus (units of inverse pressure), and μ is Poison's Ratio (dimensionless). As shown in Fig. 19.3, strain is in the direction of stress for the diagonal terms (tension stress leads to expansion) and in the opposite direction for the off-diagonal terms. In the general case, we would need to consider shear stress and shear strain. However, heat induced stress has no shear component and an isotropic material will not couple compressive stress into shear stress. A poorly

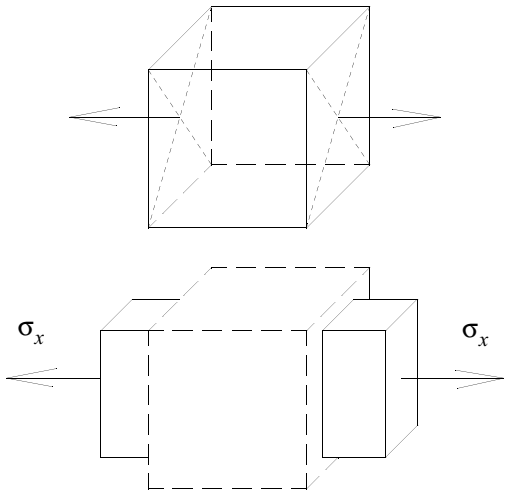


Fig. 19.3. Elastic distortion of a cube by horizontal stress. The cube elongates in the horizontal direction according to $\varepsilon_x = \sigma_x / E$ and compresses in the orthogonal directions by $\varepsilon_y = \xi_z = \mu\sigma_x / E$.

designed support can potentially apply strong forces that may induce shear effects but, we shall consider only well designed mounting that does not apply significant forces onto the part. We may, therefore, simplify the development by neglecting shear effects. If we know strains we may calculate the imposed stress

$$\begin{bmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{xy} \\ \tau xz \\ \tau_{yz} \end{bmatrix} = \frac{E}{(1+\mu)(1-2\mu)} \begin{bmatrix} 1-\mu & \mu & \mu & 0 & 0 & 0 \\ \mu & 1-\mu & \mu & 0 & 0 & 0 \\ \mu & \mu & 1-\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1-\mu}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1-\mu}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1-\mu}{2} \end{bmatrix} \begin{bmatrix} \varepsilon_x \\ \varepsilon_x \\ \varepsilon_z \\ \gamma_{xy} \\ \gamma_{xz} \\ \gamma_{yz} \end{bmatrix}, \tag{19.36}$$

$$\boldsymbol{\sigma} = \mathbf{D}\boldsymbol{\varepsilon}, \tag{19.37}$$

where **D** is the stress/strain or constitutive matrix

$$\mathbf{D} = \frac{E}{(1+\mu)(1-2\mu)}\begin{bmatrix} 1-\mu & \mu & \mu & 0 & 0 & 0 \\ \mu & 1-\mu & \mu & 0 & 0 & 0 \\ \mu & \mu & 1-\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1-\mu}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1-\mu}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1-\mu}{2} \end{bmatrix}. \tag{19.38}$$

We may compute the strain energy for this section of the solid by

$$\text{incremental energy per unit volume } \Delta U = \boldsymbol{\varepsilon}^t \mathrm{D}\boldsymbol{\varepsilon}. \tag{19.39}$$

The total strain energy is

$$U = \iiint \boldsymbol{\varepsilon}^t \mathrm{D}\boldsymbol{\varepsilon} dx dy dz. \tag{19.40}$$

Given some initial strain distribution, the system will relax by reducing the strain energy while maintaining the total strain to be constant:

$$\iiint \boldsymbol{\varepsilon} dx dy dz = \text{constant}. \tag{19.41}$$

19.3.1 Governing Equations Of Strain

For a body subject to distortions, the three components of displacement may be represented by three-dimensional functions:

$$u = u(x, y, z),\ v = v(x, y, z), \text{ and } w = w(x, y, z). \tag{19.42}$$

The strain-displacement relations are:

$$\varepsilon_x = \frac{\partial u}{\partial x},\ \varepsilon_y = \frac{\partial v}{\partial y}, \text{ and } \varepsilon_z = \frac{\partial w}{\partial z}. \tag{19.43}$$

and shear components of strain-displacement are:

$$\gamma_{xy} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x},\ \gamma_{yz} = \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}, \text{ and } \gamma_{zx} = \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}. \tag{19.44}$$

We may write the six stress-displacement functions as:

$$\sigma_x = \lambda\varepsilon + 2G\varepsilon_x,\ \sigma_y = \lambda\varepsilon + 2G\varepsilon_y, \text{ and } \sigma_z = \lambda\varepsilon + 2G\varepsilon_z, \tag{19.45}$$

$$\tau_{xy} = G\gamma_{xy},\ \tau_{yz} = G\gamma_{yz}, \text{ and } \tau_{zx} = G\gamma_{zx}, \tag{19.46}$$

where

$$\varepsilon = \varepsilon_x + \varepsilon_y + \varepsilon_z,\ G = \frac{E}{2(1+\mu)}, \text{ and } \lambda = \frac{\mu E}{(1+\mu)(1-2\mu)}. \tag{19.47}$$

The three equilibrium equations are:

$$(\lambda + G)\frac{\partial\varepsilon}{\partial x} + G\nabla^2 u + P_x = 0, \tag{19.48}$$

$$(\lambda + G)\frac{\partial\varepsilon}{\partial y} + G\nabla^2 v + P_y = 0,$$

$$(\lambda + G)\frac{\partial\varepsilon}{\partial z} + G\nabla^2 w + P_z = 0,$$

where P_x, P_y, and P_z are applied pressures. We may write the effect of temperature as

$$\begin{bmatrix}\varepsilon_x\\ \varepsilon_y\\ \varepsilon_z\end{bmatrix} = \frac{1}{E}\begin{bmatrix}1 & -\mu & -\mu\\ -\mu & 1 & -\mu\\ -\mu & -\mu & 1\end{bmatrix}\begin{bmatrix}\sigma_x\\ \sigma_y\\ \sigma_z\end{bmatrix} + \begin{bmatrix}\alpha T\\ \alpha T\\ \alpha T\end{bmatrix}. \tag{19.49}$$

The temperature gradients do not directly drive shear strain

$$\begin{bmatrix}\sigma_x\\ \sigma_y\\ \sigma_z\end{bmatrix} = \frac{E}{(1+\mu)(1-2\mu)}\begin{bmatrix}1-\mu & \mu & \mu\\ \mu & 1-\mu & \mu\\ \mu & \mu & 1-\mu\end{bmatrix}\begin{bmatrix}\varepsilon_x - \alpha T\\ \varepsilon_y - \alpha T\\ \varepsilon_z - \alpha T\end{bmatrix}. \tag{19.50}$$

Expanding Eq. (19.50) we have the stresses induced by strains and temperatures

$$\sigma_x = \frac{E}{(1+\mu)(1-2\mu)}[(1-\mu)\varepsilon_x + \mu\varepsilon_y + \mu\varepsilon_z] - \frac{E\alpha T}{(1-2\mu)}, \tag{19.51}$$

$$\sigma_y = \frac{E}{(1+\mu)(1-2\mu)}[\mu\varepsilon_x + (1-\mu)\varepsilon_y + \mu\varepsilon_z] - \frac{E\alpha T}{(1-2\mu)}, \tag{19.52}$$

$$\sigma_z = \frac{E}{(1+\mu)(1-2\mu)}[\mu\varepsilon_x + \mu\varepsilon_y + (1-\mu)\varepsilon_z] - \frac{E\alpha T}{(1-2\mu)}. \tag{19.53}$$

The stresses, shears, and pressures satisfy:

$$\frac{\partial\sigma_x}{\partial x} + \frac{\partial\tau_{yx}}{\partial y} + \frac{\partial\tau_{zx}}{\partial z} + P_x = 0, \tag{19.54}$$

$$\frac{\partial\sigma_y}{\partial y} + \frac{\partial\tau_{zy}}{\partial z} + \frac{\partial\tau_{xy}}{\partial x} + P_y = 0,$$

$$\frac{\partial\sigma_z}{\partial z} + \frac{\partial\tau_{xz}}{\partial x} + \frac{\partial\tau_{yz}}{\partial y} + P_z = 0.$$

where P_x, P_y, and P_z are external forces. Considering only thermal effects, these forces are zero. Note that we may consider the effect of thermal gradients to be equivalent to mechanical external forces as discussed below. The shears may be written in terms of derivatives of displacement:

$$\tau_{xy} = \tau_{yx} = \frac{E}{2(1+\mu)}\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right), \tag{19.55}$$

$$\tau_{xz} = \tau_{zx} = \frac{E}{2(1+\mu)}\left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right),$$

$$\tau_{yz} = \tau_{zy} = \frac{E}{2(1+\mu)}\left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right).$$

Expanding Eq. (19.51) we have the x-component of stress

$$\sigma_x = \frac{E}{(1+\mu)(1-2\mu)}\left[(1-\mu)\frac{\partial u}{\partial x} + \mu\frac{\partial v}{\partial y} + \mu\frac{\partial w}{\partial z}\right] - \frac{E\alpha T}{(1-2\mu)}. \tag{19.56}$$

Differentiating Eq. (19.56) we have

$$\frac{\partial\sigma_x}{\partial x} = \frac{E}{(1+\mu)(1-2\mu)}\left[(1-\mu)\frac{\partial^2 u}{\partial x^2} + \mu\frac{\partial^2 v}{\partial x\partial y} + \mu\frac{\partial^2 w}{\partial x\partial z}\right] - \frac{E\alpha}{(1-2\mu)}\frac{\partial T}{\partial x}. \tag{19.57}$$

The other two terms of Eq. (19.54) are

$$\frac{\partial \tau_{xy}}{\partial y} = \frac{E}{2(1+\mu)}\left(\frac{\partial^2 v}{\partial y \partial x} + \frac{\partial^2 u}{\partial y^2}\right), \tag{19.58}$$

$$\frac{\partial \tau_{xz}}{\partial z} = \frac{E}{2(1+\mu)}\left(\frac{\partial^2 u}{\partial z^2} + \frac{\partial^2 w}{\partial z \partial x}\right). \tag{19.59}$$

The equilibrium equation for the x-direction is, therefore,

$$\begin{aligned} &\frac{E}{(1+\mu)(1-2\mu)}\left[(1-\mu)\frac{\partial^2 u}{\partial x^2} + \mu\frac{\partial^2 v}{\partial x \partial y} + \mu\frac{\partial^2 w}{\partial x \partial z}\right] \\ &+ \frac{E}{2(1+\mu)}\left(\frac{\partial^2 v}{\partial y \partial x} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} + \frac{\partial^2 w}{\partial z \partial x}\right) + P_x - \frac{E\alpha}{(1-2\mu)}\frac{\partial T}{\partial x} = 0 . \end{aligned} \tag{19.60}$$

The other two equations for y- and z-directions are found by cyclically by permuting x, y, and z and u, v, and w. We may assume

$$\frac{\partial}{\partial x}\frac{\partial}{\partial y} = \frac{\partial}{\partial y}\frac{\partial}{\partial x}$$

to simplify Eq. (19.60)

$$\begin{aligned} &\frac{E(1-\mu)}{(1+\mu)(1-2\mu)}\frac{\partial^2 u}{\partial x^2} + \frac{E}{2(1+\mu)(1-2\mu)}\left(\frac{\partial^2 v}{\partial x \partial y} + \frac{\partial^2 w}{\partial x \partial z}\right) \\ &+ \frac{E}{2(1+\mu)}\left(\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) + P_x - \frac{E\alpha}{(1-2\mu)}\frac{\partial T}{\partial x} = 0 , \end{aligned} \tag{19.61}$$

(cyclical for x, y, z and u, v, w). The thermal effects may be considered to be effectively the same as mechanical pressures, such as those included in the equilibrium equations, Eq. (19.54). The thermal equivalent to mechanical pressures are

$$-\frac{E\alpha}{1-2\mu}\frac{\partial T}{\partial x}, \text{ cylical for } x, y, z \tag{19.62}$$

We may convert the equations above to two-dimensional form by specifying $\varepsilon_z = 0$, $\tau_{yz} = 0$, and $\tau_{zx} = 0$. $\varepsilon_z = 0$ implies $\sigma_z = \mu(\sigma_x + \sigma_y) - E\alpha T$. The two-dimensional equations are:

$$\begin{bmatrix} \varepsilon_x \\ \varepsilon_y \end{bmatrix} = \frac{1}{E}\begin{bmatrix} 1 & -\mu \\ -\mu & 1 \end{bmatrix}\begin{bmatrix} \sigma_x \\ \sigma_y \end{bmatrix} + \begin{bmatrix} \alpha T \\ \alpha T \end{bmatrix}, \tag{19.63}$$

$$\begin{bmatrix}\sigma_x\\ \sigma_y\end{bmatrix} = \frac{E}{(1+\mu)(1-\mu)}\begin{bmatrix}1-\mu & \mu\\ \mu & 1-\mu\end{bmatrix}\begin{bmatrix}\varepsilon_x-\alpha T\\ \varepsilon_y-\alpha T\end{bmatrix}. \tag{19.64}$$

Expanding Eq. (19.50) we have the stresses induced by strains and temperatures

$$\sigma_x = \frac{E}{(1+\mu)(1-\mu)}\left[(1-\mu)\varepsilon_x+\mu\varepsilon_y\right]-\frac{E\alpha T}{(1-\mu)}, \tag{19.65}$$

$$\sigma_y = \frac{E}{(1+\mu)(1-\mu)}\left[\mu\varepsilon_x+(1-\mu)\varepsilon_y\right]-\frac{E\alpha T}{(1-\mu)}. \tag{19.66}$$

The stresses, shears, and pressures satisfy:

$$\frac{\partial\sigma_x}{\partial x}+\frac{\partial\tau_{yx}}{\partial y}+P_x = 0 \quad , \tag{19.67}$$

$$\frac{\partial\sigma_y}{\partial y}+\frac{\partial\tau_{xy}}{\partial x}+P_y = 0 \quad , \tag{19.68}$$

where P_x, and P_y, are external mechanical forces that are zero for strictly thermal effects.

$$\frac{E}{(1+\mu)(1-\mu)}\frac{\partial^2 u}{\partial x^2}+\frac{E}{2(1-\mu)}\frac{\partial^2 v}{\partial x\partial y}+\frac{E}{2(1+\mu)}\frac{\partial^2 u}{\partial y^2}+P_x-\frac{E\alpha}{(1-\mu)}\frac{\partial T}{\partial x} = 0, \tag{19.69}$$

$$\frac{E}{(1+\mu)(1-\mu)}\frac{\partial^2 v}{\partial y^2}+\frac{E}{2(1-\mu)}\frac{\partial^2 u}{\partial y\partial x}+\frac{E}{2(1+\mu)}\frac{\partial^2 v}{\partial x^2}+P_x-\frac{E\alpha}{(1-\mu)}\frac{\partial T}{\partial y} = 0. \tag{19.70}$$

Simplifying the above equations and assuming no external forces,

$$\frac{\partial^2 u}{\partial x^2}+\frac{(1-\mu)}{2}\frac{\partial^2 u}{\partial y^2}+\frac{(1+\mu)}{2}\frac{\partial^2 v}{\partial y\partial x}-\alpha(1+\mu)\frac{\partial T}{\partial x} = 0, \tag{19.71}$$

$$\frac{\partial^2 v}{\partial y^2}+\frac{(1-\mu)}{2}\frac{\partial^2 v}{\partial x^2}+\frac{(1+\mu)}{2}\frac{\partial^2 u}{\partial x\partial y}-\alpha(1+\mu)\frac{\partial T}{\partial y} = 0. \tag{19.72}$$

These equations may be evaluated numerically in terms of nearest neighbors of a point $(0, 0)$.

The three-dimensional equilibrium equations are of the form:

$$\frac{E(1-\mu)}{(1+\mu)(1-2\mu)}\frac{\partial^2 u}{\partial x^2}+\frac{E}{2(1+\mu)(1-2\mu)}\left(\frac{\partial^2 v}{\partial x\partial y}+\frac{\partial^2 w}{\partial x\partial z}\right) \tag{19.73}$$
$$+\frac{E}{2(1+\mu)}\left(\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right)+P_x-\frac{E\alpha}{(1-2\mu)}\frac{\partial T}{\partial x}=0\,,$$

(cyclical for x, y, z).

Defining four coefficients

$$C_1\frac{\partial^2 u}{\partial x^2}+C_2\left(\frac{\partial^2 v}{\partial x\partial y}+\frac{\partial^2 w}{\partial x\partial z}\right)+C_3\left(\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right)+P_x+C_4\frac{\partial T}{\partial x}=0\,, \tag{19.74}$$

$$C_1=\frac{E(1-\mu)}{(1+\mu)(1-2\mu)},\ C_2=\frac{E}{2(1+\mu)(1-2\mu)},\ C_3=\frac{E}{2(1+\mu)},\ \text{and}\ C_4=-\frac{E\alpha}{(1-2\mu)}. \tag{19.75}$$

19.3.1.1 Finite Difference Solution, 3D

The derivatives may be evaluated by finite difference equations. The ensemble of 27 points in the cube of the local neighborhood may be identified by relative indices (i, j, k), where $i, j, k = -1, 0, 1$. The C-coefficients are evaluated at (0,0,0):

$$\frac{\partial^2 u}{\partial x^2}\approx\frac{u(1,0,0)+u(-1,0,0)-2u(0,0,0)}{\Delta x^2}, \tag{19.76}$$

$$\frac{\partial^2 v}{\partial x\partial y}\approx\frac{v(1,1,0)-v(1,-1,0)+v(-1,-1,0)-v(-1,1,0)}{4\Delta x\Delta y}, \tag{19.77}$$

$$\frac{\partial^2 w}{\partial x\partial z}\approx\frac{w(1,0,1)-w(1,0,-1)+w(-1,0,-1)-w(-1,0,1)}{4\Delta x\Delta z}, \tag{19.78}$$

$$\frac{\partial^2 u}{\partial y^2}\approx\frac{u(0,1,0)+u(0,-1,0)-2u(0,0,0)}{\Delta y^2}, \tag{19.79}$$

$$\frac{\partial^2 u}{\partial z^2}\approx\frac{u(0,0,1)+u(0,0,-1)-2u(0,0,0)}{\Delta z^2}, \tag{19.80}$$

$$\frac{\partial T}{\partial x}\approx\frac{T(1,0,0)-T(-1,0,0)}{2\Delta x}. \tag{19.81}$$

The solution for $u(0, 0, 0)$ that forces compliance with the equilibrium equation is

$$u(0,0,0) = \frac{1}{2C_1 + 4C_3}\left[C_1\frac{u(1,0,0)+u(-1,0,0)}{\Delta x^2}\right. \tag{19.82}$$
$$+ C_2\frac{v(1,1,0)-v(1,-1,0)+v(-1,-1,0)-v(-1,1,0)}{4\Delta x\Delta y}$$
$$+ C_2\frac{w(1,0,1)-w(1,0,-1)+w(-1,0,-1)-w(-1,0,1)}{4\Delta x\Delta z} + C_3\frac{u(0,1,0)+u(0,-1,0)}{\Delta y^2}$$
$$\left. + C_3\frac{u(0,0,1)+u(0,0,-1)}{\Delta z^2} + P_x'(0,0,0)\right].$$

We use the generalized pressure P_x' to include mechanical and temperature gradient effects.

$$P_x' = P_x(0,0,0) + C_4\frac{T(1,0,0)-T(-1,0,0)}{2\Delta x}, \tag{19.83}$$

(similarly for $v(0,0,0)$ and $w(0,0,0)$ with cyclical permutation of x, y, and z and u, v, and w).

$$v(0,0,0) = \frac{1}{2C_1 + 4C_3}\left[C_1\frac{v(0,1,0)+v(0,-1,0)}{\Delta y^2}\right. \tag{19.84}$$
$$+ C_2\frac{w(0,1,1)-w(0,1,-1)+w(0,-1,-1)-w(0,-1,1)}{4\Delta y\Delta z}$$
$$+ C_2\frac{u(1,1,0)-u(-1,1,0)+u(-1,-1,0)-u(1,-1,0)}{4\Delta y\Delta x} + C_3\frac{v(0,0,1)+v(0,0,-1)}{\Delta z^2}$$
$$\left. + C_3\frac{v(1,0,0)+v(-1,0,0)}{\Delta x^2} + P_y'(0,0,0)\right],$$

$$P_y' = P_y(0,0,0) + C_4\frac{T(0,1,0)-T(0,-1,0)}{2\Delta y}, \tag{19.85}$$

$$w(0,0,0) = \frac{1}{2C_1 + 4C_3}\left[C_1\frac{w(0,0,1)+w(0,0,-1)}{\Delta z^2}\right. \tag{19.86}$$
$$+ C_2\frac{u(1,0,1)-u(-1,0,1)+u(-1,0,-1)-u(1,0,-1)}{4\Delta z\Delta x}$$
$$+ C_2\frac{v(0,1,1)-v(0,-1,1)+v(0,-1,-1)-v(0,1,-1)}{4\Delta z\Delta y} + C_3\frac{w(1,0,0)+w(-1,0,0)}{\Delta x^2}$$
$$\left. + C_3\frac{w(0,1,0)+w(0,-1,0)}{\Delta y^2} + P_z'(0,0,0)\right],$$

$$P_z' = P_z(0,0,0) + C_4\frac{T(0,0,1)-T(0,0,-1)}{2\Delta z}. \tag{19.87}$$

As we will represent the volume by x,y planes along the z-axis, we resort the terms according to z-value.

$$u(0,0,0) = \frac{1}{2C_1+4C_3}\Bigg[C_2\frac{w(-1,0,-1)-w(1,0,-1)}{4\Delta x\Delta z}+C_3\frac{u(0,0,-1)}{\Delta z^2} \tag{19.88}$$
$$+C_2\frac{v(1,1,0)-v(-1,1,0)}{4\Delta x\Delta y}+C_3\frac{u(0,1,0)}{\Delta y^2}+C_1\frac{u(1,0,0)+u(-1,0,0)}{\Delta x^2}+P_x'(0,0,0)$$
$$+C_2\frac{v(-1,-1,0)+v(1,-1,0)}{4\Delta x\Delta y}+C_3\frac{u(0,-1,0)}{\Delta y^2}$$
$$+C_2\frac{w(1,0,1)-w(-1,0,1)}{4\Delta x\Delta z}+C_3\frac{u(0,0,1)}{\Delta z^2}\Bigg],$$

$$v(0,0,0) = \frac{1}{2C_1+4C_3}\Bigg[-C_2\frac{w(0,1,-1)}{4\Delta y\Delta z}+C_3\frac{v(0,0,-1)}{\Delta z^2}+C_2\frac{w(0,-1,-1)}{4\Delta y\Delta z} \tag{19.89}$$
$$+C_1\frac{v(0,1,0)}{\Delta y^2}+C_2\frac{u(1,0,0)-u(-1,1,0)}{4\Delta y\Delta x}+C_3\frac{v(1,0,0)+v(-1,0,0)}{\Delta x^2}+P_y'(0,0,0)+C_1\frac{v(0,-1,0)}{\Delta y^2}$$
$$+C_2\frac{u(-1,-1,0)-u(1,-1,0)}{4\Delta y\Delta x}$$
$$+C_2\frac{w(0,1,1)}{4\Delta y\Delta z}+C_3\frac{v(0,0,1)}{\Delta z^2}-C_2\frac{w(0,-1,1)}{4\Delta y\Delta z}\Bigg],$$

$$w(0,0,0) = \frac{1}{2C_1+4C_3}\Bigg[-C_2\frac{v(0,1,-1)}{4\Delta z\Delta y}+C_2\frac{u(-1,0,-1)-u(1,0,-1)}{4\Delta z\Delta x}+C_1\frac{w(0,0,-1)}{\Delta z^2} \tag{19.90}$$
$$+C_2\frac{v(0,-1,-1)}{4\Delta z\Delta y}$$
$$+C_3\frac{w(0,1,0)}{\Delta y^2}+C_3\frac{w(1,0,0)+w(-1,0,0)}{\Delta x^2}+P_z'(0,0,0)+C_3\frac{w(0,-1,0)}{\Delta y^2}$$
$$+C_2\frac{v(0,1,1)}{4\Delta z\Delta y}+C_2\frac{u(1,0,1)-u(-1,0,1)}{4\Delta z\Delta x}+C_1\frac{w(0,0,1)}{\Delta z^2}-C_2\frac{v(0,-1,1)}{4\Delta z\Delta y}\Bigg].$$

The variable material terms contain only first derivatives and are independent of the center point.

19.3.2 Thermal Stress in a Plane: Two-dimensional Solution

The two-dimensional solution may be derived as a special case of the three-dimensional solution, or directly from the two-dimensional form of the fundamental equations:

$$\text{Hooke's Law, } \begin{bmatrix}\sigma_x\\ \sigma_y\\ \tau_{xy}\end{bmatrix} = \frac{E}{(1+\mu)(1-\mu)}\begin{bmatrix}1 & \mu & 0\\ \mu & 1 & 0\\ 0 & 0 & \frac{(1-\mu)}{2}\end{bmatrix}\begin{bmatrix}\varepsilon_x\\ \varepsilon_y\\ \gamma_{xy}\end{bmatrix}. \tag{19.91}$$

$$\text{Hooke's Law with thermal driving forces, } \begin{bmatrix}\sigma_x\\ \sigma_y\\ \tau_{xy}\end{bmatrix} = \frac{E}{(1+\mu)(1-\mu)}\begin{bmatrix}1 & \mu & 0\\ \mu & 1 & 0\\ 0 & 0 & \frac{(1-\mu)}{2}\end{bmatrix}\begin{bmatrix}\varepsilon_x - \alpha T\\ \varepsilon_y - \alpha T\\ \gamma_{xy}\end{bmatrix}. \quad (19.92)$$

Strains induced by local stresses and local temperatures are found by inverse Hooke's Law:

$$\text{inverse Hooke's law, } \begin{bmatrix}\varepsilon_x\\ \varepsilon_y\\ \gamma_{xy}\end{bmatrix} = \frac{1}{E}\begin{bmatrix}1 & -\mu & 0\\ -\mu & 1 & 0\\ 0 & 0 & 2(1+\mu)\end{bmatrix}\begin{bmatrix}\sigma_x\\ \sigma_y\\ \tau_{xy}\end{bmatrix} + \begin{bmatrix}\alpha T\\ \alpha T\\ 0\end{bmatrix}. \quad (19.93)$$

Expanding the matrix formulation into two equations we have

$$\sigma_x = \frac{E}{(1+\mu)(1-\mu)}(\varepsilon_x + \mu\varepsilon_y) - \frac{E\alpha T}{(1-\mu)}, \quad (19.94)$$

$$\sigma_y = \frac{E}{(1+\mu)(1-\mu)}(\mu\varepsilon_x + \varepsilon_y) - \frac{E\alpha T}{(1-\mu)}, \quad (19.95)$$

$$\tau_{xy} = \frac{E}{2(1+\mu)}\gamma_{xy}. \quad (19.96)$$

The equilibrium equations define the steady-state conditions (from 19.94 and 19.95) when the material has time to relax:

$$\frac{\partial \sigma_x}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} + P_x = 0 \quad , \quad (19.97)$$

$$\frac{\partial \sigma_y}{\partial y} + \frac{\partial \tau_{xy}}{\partial x} + P_y = 0 \quad . \quad (19.98)$$

The strain-displacement relations in 2D are:

$$\varepsilon_x = \frac{\partial u}{\partial x},\ \varepsilon_y = \frac{\partial v}{\partial y}, \text{ and } \gamma_{xy} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}. \quad (19.99)$$

We may write the equilibrium equations in terms of strains by replacing stress values with strain values according to Hooke's Law

$$\frac{\partial \sigma_x}{\partial x} = \frac{E}{(1+\mu)(1-\mu)}\left(\frac{\partial^2 u}{\partial x^2} + \mu\frac{\partial^2 v}{\partial x \partial y}\right) - \frac{E\alpha}{(1-\mu)}\frac{\partial T}{\partial x}, \tag{19.100}$$

$$\frac{\partial \tau_{xy}}{\partial y} = \frac{E}{2(1+\mu)}\left(\frac{\partial^2 v}{\partial y \partial x} + \frac{\partial^2 u}{\partial y^2}\right), \tag{19.101}$$

From Eqs. (19.97), (19.100), and (19.101)

$$\frac{E}{(1+\mu)(1-\mu)}\left(\frac{\partial^2 u}{\partial x^2} + \mu\frac{\partial^2 v}{\partial x \partial y}\right) + \frac{E}{2(1+\mu)}\left(\frac{\partial^2 v}{\partial y \partial x} + \frac{\partial^2 u}{\partial y^2}\right) - \frac{E\alpha}{(1-\mu)}\frac{\partial T}{\partial x} = 0, \tag{19.102}$$

(cyclical for x, y and u, v).

Again assuming

$$\frac{\partial}{\partial x}\frac{\partial}{\partial y} = \frac{\partial}{\partial y}\frac{\partial}{\partial x},$$

$$\frac{E}{(1+\mu)(1-\mu)}\frac{\partial^2 u}{\partial x^2} + \frac{E}{2(1-\mu)}\frac{\partial^2 v}{\partial x \partial y} + \frac{E}{2(1+\mu)}\frac{\partial^2 u}{\partial y^2} - \frac{E\alpha}{(1-\mu)}\frac{\partial T}{\partial x} = 0, \tag{19.103}$$

(cyclical for x, y and u, v).

Writing Eq. (19.103) in terms of three C-coefficients, we have

$$C_1\frac{\partial^2 u}{\partial x^2} + (\mu C_1 + C_2)\frac{\partial^2 v}{\partial x \partial y} + C_2\frac{\partial^2 u}{\partial y^2} + P_x + C_3\frac{\partial T}{\partial x} = 0, \tag{19.104}$$

$$C_1 = \frac{E}{(1+\mu)(1-\mu)},\ C_2 = \frac{E}{2(1+\mu)},\ C_3 = -\frac{E\alpha}{(1-2\mu)} \text{ and } P_x = 0. \tag{19.105}$$

19.3.2.1 Finite Difference Solution, 2D

The ensemble of nine points in the square of the local neighborhood may be identified by relative indices such that (i, j), where $i, j = -1, 0, 1$. The C-coefficients are evaluated at (0,0)

$$\frac{\partial^2 u}{\partial x^2} \approx \frac{u(1,0) + u(-1,0) - 2u(0,0)}{\Delta x^2}, \tag{19.106}$$

$$\frac{\partial^2 v}{\partial x \partial y} \approx \frac{v(1,1) - v(1,-1) + v(-1,-1) - v(-1,1)}{4\Delta x \Delta y}, \tag{19.107}$$

$$\frac{\partial^2 u}{\partial y^2} \approx \frac{u(0,1)+u(0,-1)-2u(0,0)}{\Delta y^2}, \tag{19.108}$$

$$\frac{\partial T}{\partial x} \approx \frac{T(1,0)-T(-1,0)}{2\Delta x}, \tag{19.109}$$

(cyclical for x, y).
The finite difference solutions are

$$u(0,0) = \frac{1}{2\left(\frac{C_1}{\Delta x^2}+\frac{C_2}{\Delta y^2}\right)}\left[C_1\frac{u(1,0)+u(-1,0)}{\Delta x^2}+C_2\frac{u(0,1)+u(0,-1)}{\Delta y^2}\right. \tag{19.110}$$
$$\left.+\frac{\mu C_1+C_2}{4\Delta x\Delta y}\left[v(1,1)-v(1,-1)+v(-1,-1)-v(-1,1)\right]+P_x+C_3\frac{T(1,0)-T(-1,0)}{2\Delta x}\right],$$

Reorganizing by y-values

$$u(0,0) = \frac{1}{2\left(\frac{C_1}{\Delta x^2}+\frac{C_2}{\Delta y^2}\right)}\left[-\frac{\mu C_1+C_2}{4\Delta x\Delta y}v(-1,1)+\frac{C_2}{\Delta y^2}u(0,1)+\frac{\mu C_1+C_2}{4\Delta x\Delta y}v(1,1)\right. \tag{19.111}$$
$$+\frac{C_1}{\Delta x^2}u(-1,0)+P_x'(0,0)+\frac{C_1}{\Delta x^2}u(1,0)$$
$$\left.+\frac{\mu C_1+C_2}{4\Delta x\Delta y}v(-1,-1)+\frac{C_2}{\Delta y^2}u(0,-1)-\frac{\mu C_1+C_2}{4\Delta x\Delta y}v(1,-1)\right],$$

$$P_x'(0,0) = P_x(0,0)+\frac{C_3}{2\Delta x}(T(1,0)-T(-1,0)), \tag{19.112}$$

$$v(0,0) = \frac{1}{2\left(\frac{C_1}{\Delta y^2}+\frac{C_2}{\Delta x^2}\right)}\left[C_1\frac{v(0,1)+v(0,-1)}{\Delta y^2}+C_2\frac{v(1,0)+v(-1,0)}{\Delta x^2}\right. \tag{19.113}$$

$$\left.+\frac{\mu C_1+C_2}{4\Delta y\Delta x}\left[u(1,1)-u(-1,1)+u(-1,-1)-u(1,-1)\right]+P_y+C_3\frac{T(0,1)-T(0,-1)}{2\Delta y}\right],$$

$$u(0,0) = \frac{1}{2\left(\frac{C_1}{\Delta x^2}+\frac{C_2}{\Delta y^2}\right)}\left[C_1\frac{u(1,0)+u(-1,0)}{\Delta x^2}+C_2\frac{u(0,1)+u(0,-1)}{\Delta y^2}\right. \tag{19.114}$$

$$\left.+\frac{\mu C_1+C_2}{4\Delta x\Delta y}\left[v(1,1)-v(1,-1)+v(-1,-1)-v(-1,1)\right]+P_x+C_3\frac{T(1,0)-T(-1,0)}{2\Delta x}\right],$$

Reorganizing by y-values

$$v(0,0) = \frac{1}{2\left(\frac{C_1}{\Delta y^2}+\frac{C_2}{\Delta x^2}\right)}\left[-\frac{\mu C_1+C_2}{4\Delta y\Delta x}u(-1,1)+\frac{C_1}{\Delta y^2}v(0,1)+\frac{\mu C_1+C_2}{4\Delta y\Delta x}u(1,1)\right. \tag{19.115}$$

$$+\frac{C_2}{\Delta x^2}v(-1,0)+P_y'(0,0)+\frac{C_2}{\Delta x^2}v(1,0)$$

$$\left.+\frac{\mu C_1+C_2}{4\Delta y\Delta x}u(-1,-1)+\frac{C_1}{\Delta y^2}v(0,-1)-\frac{\mu C_1+C_2}{4\Delta y\Delta x}u(1,-1)\right],$$

$$P_y'(0,0) = P_y(0,0)+\frac{C_3}{2\Delta y}(T(0,1)-T(0,-1)), \tag{19.116}$$

$$u(0,0) = \{D_{u1}[v(1,1)-v(1,-1)+v(-1,-1)-v(-1,1)]+D_{u2}[u(0,1)+u(0,-1)] \tag{19.117}$$
$$+D_{u3}[u(1,0)+u(-1,0)]+D_{u4}P_x'\},$$

$$D_{u1} = \frac{\frac{\mu C_1 + C_2}{4\Delta x \Delta y}}{2\left(\frac{C_1}{\Delta x^2} + \frac{C_2}{\Delta y^2}\right)}, D_{u2} = \frac{\frac{C_2}{\Delta y^2}}{2\left(\frac{C_1}{\Delta x^2} + \frac{C_2}{\Delta y^2}\right)}, D_{u3} = \frac{\frac{C_1}{\Delta x^2}}{2\left(\frac{C_1}{\Delta x^2} + \frac{C_2}{\Delta y^2}\right)}, D_{u4} = \frac{\frac{C_3}{2\Delta x}}{2\left(\frac{C_1}{\Delta x^2} + \frac{C_2}{\Delta y^2}\right)}, \quad (19.118)$$

$$v(0,0) = \{D_{v1}[u(1,1) - u(1,-1) + u(-1,-1) - u(-1,1)] + D_{v2}[v(1,0) + v(-1,0)] \quad (19.119)$$
$$+ D_{v3}[v(0,1) + v(-0,-1)] + D_{v4}P_x'\},$$

$$D_{v1} = \frac{\frac{\mu C_1 + C_2}{4\Delta x \Delta y}}{2\left(\frac{C_1}{\Delta y^2} + \frac{C_2}{\Delta x^2}\right)}, D_{v2} = \frac{\frac{C_2}{\Delta x^2}}{2\left(\frac{C_1}{\Delta y^2} + \frac{C_2}{\Delta x^2}\right)}, D_{v3} = \frac{\frac{C_1}{\Delta y^2}}{2\left(\frac{C_1}{\Delta y^2} + \frac{C_2}{\Delta x^2}\right)}, D_{v4} = \frac{\frac{C_3}{2\Delta y}}{2\left(\frac{C_1}{\Delta y^2} + \frac{C_2}{\Delta x^2}\right)}. \quad (19.120)$$

The effective forces due to thermal gradients are independent of displacements and may be computed outside the scan calculations needed to compute the displacements.

19.3.3 Green's Function Method

We may find the Fourier transform of the Green's Function by solving the differential equations in the Fourier domain. The Fourier transform of the Green's Function may be used as the transfer function for stress operating on the Fourier transform of the temperature distribution.

19.3.3.1 Green's Function Solution of 2D Thermal Stress

$$\frac{1}{(1+\mu)(1-\mu)}\frac{\partial^2 u}{\partial x^2} + \frac{1}{2(1-\mu)}\frac{\partial^2 v}{\partial x \partial y} + \frac{1}{2(1+\mu)}\frac{\partial^2 u}{\partial y^2} - \frac{\alpha}{(1-\mu)}\frac{\partial T}{\partial x} = 0, \quad (19.121)$$

$$\frac{1}{(1+\mu)(1-\mu)}\frac{\partial^2 v}{\partial y^2} + \frac{1}{2(1-\mu)}\frac{\partial^2 u}{\partial y \partial x} + \frac{1}{2(1+\mu)}\frac{\partial^2 v}{\partial x^2} - \frac{\alpha}{(1-\mu)}\frac{\partial T}{\partial y} = 0, \quad (19.122)$$

$$\frac{\partial^2 u}{\partial x^2} + \frac{1+\mu}{2}\frac{\partial^2 v}{\partial x \partial y} + \frac{1-\mu}{2}\frac{\partial^2 u}{\partial y^2} = \alpha(1+\mu)\frac{\partial T}{\partial x}, \quad (19.123)$$

$$\frac{\partial^2 v}{\partial y^2} + \frac{1+\mu}{2}\frac{\partial^2 u}{\partial y \partial x} + \frac{1-\mu}{2}\frac{\partial^2 v}{\partial x^2} = \alpha(1+\mu)\frac{\partial T}{\partial y}. \quad (19.124)$$

Taking the Fourier transform of both equations we have

$$\left(\xi^2 + \eta^2 \frac{1-\mu}{2}\right) U(\xi,\eta) + \xi\eta \frac{1+\mu}{2} V(\xi,\eta) = -\frac{j\alpha(1+\mu)}{2\pi} \tilde{T}(\xi,\eta)\xi, \tag{19.125}$$

$$\left(\eta^2 + \xi^2 \frac{1-\mu}{2}\right) V(\xi,\eta) + \eta\xi \frac{1+\mu}{2} U(\xi,\eta) = -\frac{j\alpha(1+\mu)}{2\pi} \tilde{T}(\xi,\eta)\eta. \tag{19.126}$$

Collecting variables we have

$$aU(\xi,\eta) + bV(\xi,\eta) = c, \tag{19.127}$$

$$dV(\xi,\eta) + bU(\xi,\eta) = e, \tag{19.128}$$

where the coefficients are

$$a = \xi^2 + \eta^2 \frac{1-\mu}{2},\ b = \xi\eta \frac{1+\mu}{2},\ c = -\frac{j\alpha(1+\mu)}{2\pi} \tilde{T}(\xi,\eta)\xi, \tag{19.129}$$

$d = \eta^2 + \xi^2 \frac{1-\mu}{2}$, and $e = -\frac{j\alpha(1+\mu)}{2\pi} \tilde{T}(\xi,\eta)\eta$.

The solution to the two coupled equations is

$$U(\xi,\eta) = \frac{cd - be}{ad - b^2},\ V(\xi,\eta) = \frac{ae - cb}{ad - b^2}. \tag{19.130}$$

We now have the transfer functions for displacement operating on the Fourier transform of the temperature distribution

$$U(\xi,\eta) = -\frac{j\alpha(1+\mu)\xi}{2\pi(\xi^2 + \eta^2)} \tilde{T}(\xi,\eta),\ V(\xi,\eta) = -\frac{j\alpha(1+\mu)\eta}{2\pi(\xi^2 + \eta^2)} \tilde{T}(\xi,\eta). \tag{19.131}$$

The strain-displacement equations (Eq. (19.99)) may also be Fourier transformed. Yielding the transfer functions for strain to operate on the Fourier transform of the temperature distribution

$$\tilde{\varepsilon}_x(\xi,\eta) = j2\pi\xi U(\xi,\eta) = \frac{\alpha(1+\mu)\xi^2}{\xi^2 + \eta^2} \tilde{T}(\xi,\eta), \qquad \tilde{\varepsilon}_x(0,0) = 0, \tag{19.132}$$

$$\tilde{\varepsilon}_y(\xi,\eta) = j2\pi\eta V(\xi,\eta) = \frac{\alpha(1+\mu)\eta^2}{\xi^2 + \eta^2} \tilde{T}(\xi,\eta), \qquad \tilde{\varepsilon}_y(0,0) = 0, \tag{19.133}$$

$$\tilde{\gamma}_{xy}(\xi, \eta) = j2\pi\eta U(\xi, \eta) + j2\pi\xi V(\xi, \eta) = \frac{2\alpha(1+\mu)\xi\eta}{\xi^2 + \eta^2}\tilde{T}(\xi, \eta), \qquad \tilde{\gamma}_{xy}(0, 0) = 0. \tag{19.134}$$

Given the transfer functions for strain we may use the frequency space form of the stress-strain equations:

$$\tilde{\sigma}_x(\xi, \eta) = \frac{E}{(1+\mu)(1-\mu)}[\tilde{\varepsilon}_x(\xi, \eta) + \mu\tilde{\varepsilon}_y(\xi, \eta)] - \frac{E\alpha\tilde{T}(\xi, \eta)}{1-\mu}, \tag{19.135}$$

$$\tilde{\sigma}_y(\xi, \eta) = \frac{E}{(1+\mu)(1-\mu)}\left[\tilde{\mu\varepsilon}_x(\xi, \eta) + \tilde{\varepsilon}_y(\xi, \eta)\right] - \frac{E\alpha\tilde{T}(\xi, \eta)}{1-\mu}, \tag{19.136}$$

$$\tilde{\tau}_{xy}(\xi, \eta) = \frac{E}{2(1+\mu)}\tilde{\gamma}_{xy}(\xi, \eta). \tag{19.137}$$

We now have the transfer functions for the components of stress to multiply the Fourier transform of the temperature distribution:

$$\tilde{\sigma}_x(\xi, \eta) = -\frac{E\alpha\eta^2}{\xi^2 + \eta^2}\tilde{T}(\xi, \eta), \qquad \tilde{\sigma}_x(0, 0) = 0, \tag{19.138}$$

$$\tilde{\sigma}_y(\xi, \eta) = -\frac{E\alpha\xi^2}{\xi^2 + \eta^2}\tilde{T}(\xi, \eta), \qquad \tilde{\sigma}_y(0, 0) = 0, \tag{19.139}$$

$$\tilde{\tau}_{xy}(\xi, \eta) = \frac{E\alpha\xi\eta}{\xi^2 + \eta^2}\tilde{T}(\xi, \eta), \qquad \tilde{\tau}_{xy}(0, 0) = 0. \tag{19.140}$$

The center point of the transfer function is not well defined. Setting $\tilde{\sigma}_x(0, 0) = 0$ makes the stress independent of the average temperature. This makes the Green's function solution more closely approximate the case of unconstrained boundaries. Note also that the stress equations are independent of Poisson's Ratio μ.

Consider a one-dimensional distribution $T(x) = T_0\exp(-x^2/\omega^2)$. The Fourier transforms of the strain components are:

$$\tilde{\sigma}_x(\xi, \eta) = -\frac{E\alpha\eta^2}{\xi^2 + \eta^2}\tilde{T}(\xi)\delta(\eta), \qquad \tilde{\sigma}_x(0, 0) = 0, \tag{19.141}$$

$$\tilde{\sigma}_y(\xi, \eta) = -\frac{E\alpha\xi^2}{\xi^2 + \eta^2}\tilde{T}(\xi)\delta(\eta), \qquad \tilde{\sigma}_y(0, 0) = 0, \tag{19.142}$$

$$\tilde{\gamma}_{xy}(\xi,\eta) = \frac{E\alpha\xi\eta}{\xi^2+\eta^2}\tilde{T}(\xi)\delta(\eta) = 0, \qquad \tilde{\gamma}_{xy}(0,0) = 0. \tag{19.143}$$

19.3.3.2 Green's Function Solution of 3D Thermal Stress

From Eq. (19.73)

$$\frac{(1-\mu)}{(1+\mu)(1-2\mu)}\frac{\partial^2 u}{\partial x^2} + \frac{1}{2(1+\mu)(1-2\mu)}\left(\frac{\partial^2 v}{\partial x\partial y} + \frac{\partial^2 w}{\partial x\partial z}\right) \tag{19.144}$$
$$+ \frac{1}{2(1+\mu)}\left(\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) - \frac{\alpha}{(1-2\mu)}\frac{\partial T}{\partial x} = 0,$$

(cyclical for x,y, z). Further

$$(1-\mu)\frac{\partial^2 u}{\partial x^2} + \frac{1}{2}\left(\frac{\partial^2 v}{\partial x\partial y} + \frac{\partial^2 w}{\partial x\partial z}\right) + \frac{1-2\mu}{2}\left(\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) = \alpha(1+\mu)\frac{\partial T}{\partial x}, \tag{19.145}$$

$$(1-\mu)\frac{\partial^2 v}{\partial y^2} + \frac{1}{2}\left(\frac{\partial^2 w}{\partial y\partial z} + \frac{\partial^2 u}{\partial y\partial x}\right) + \frac{1-2\mu}{2}\left(\frac{\partial^2 v}{\partial z^2} + \frac{\partial^2 v}{\partial x^2}\right) = \alpha(1+\mu)\frac{\partial T}{\partial y}, \tag{19.146}$$

$$(1-\mu)\frac{\partial^2 w}{\partial z^2} + \frac{1}{2}\left(\frac{\partial^2 u}{\partial z\partial x} + \frac{\partial^2 v}{\partial z\partial y}\right) + \frac{1-2\mu}{2}\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2}\right) = \alpha(1+\mu)\frac{\partial T}{\partial z}. \tag{19.147}$$

Taking the Fourier transform we have

$$(1-\mu)\xi^2 U(\xi,\eta,\omega) + \frac{1}{2}[\xi\eta V(\xi,\eta,\omega) + \xi\omega W(\xi,\eta,\omega)] \tag{19.148}$$
$$+ \frac{1-2\mu}{2}[\eta^2+\omega^2]U(\xi,\eta,\omega) = -\frac{j}{2\pi}\alpha(1+\mu)\xi\tilde{T}(\xi,\eta,\omega),$$

$$(1-\mu)\eta^2 V(\xi,\eta,\omega) + \frac{1}{2}[\eta\omega W(\xi,\eta,\omega) + \eta\xi U(\xi,\eta,\omega)] \tag{19.149}$$
$$+ \frac{1-2\mu}{2}[\omega^2+\xi^2]V(\xi,\eta,\omega) = -\frac{j}{2\pi}\alpha(1+\mu)\eta\tilde{T}(\xi,\eta,\omega),$$

$$(1-\mu)\omega^2 W(\xi,\eta,\omega) + \frac{1}{2}[\omega\xi U(\xi,\eta,\omega) + \omega\eta V(\xi,\eta,\omega)] \tag{19.150}$$
$$+ \frac{1-2\mu}{2}[\xi^2+\eta^2]W(\xi,\eta,\omega) = -\frac{j}{2\pi}\alpha(1+\mu)\omega\tilde{T}(\xi,\eta,\omega),$$

$$\left[(1-\mu)\xi^2+\frac{1-2\mu}{2}(\eta^2+\omega^2)\right]U+\frac{1}{2}\xi\eta V+\frac{1}{2}\xi\omega W=-\frac{j\alpha(1+\mu)\xi}{2\pi}\tilde{T}, \tag{19.151}$$

$$\left[(1-\mu)\eta^2+\frac{1-2\mu}{2}(\omega^2+\xi^2)\right]V+\frac{1}{2}\eta\omega W+\frac{1}{2}\eta\xi U=-\frac{j\alpha(1+\mu)\eta}{2\pi}\tilde{T}, \tag{19.152}$$

$$\left[(1-\mu)\omega^2+\frac{1-2\mu}{2}(\xi^2+\eta^2)\right]W+\frac{1}{2}\omega\xi U+\frac{1}{2}\omega\xi U=-\frac{j\alpha(1+\mu)\omega}{2\pi}\tilde{T}, \tag{19.153}$$

$$aU+bV+cW=d\tilde{T}, \tag{19.154}$$

$$eU+fV+gW=h\tilde{T}, \tag{19.155}$$

$$iU+jV+kW=l\tilde{T}, \tag{19.156}$$

$$\begin{bmatrix}U\\V\\W\end{bmatrix}=\frac{\begin{bmatrix}fk-gj & cj-bk & bg-cf\\ gi-ek & ak-ci & ce-ag\\ ej-if & bi-aj & af-be\end{bmatrix}}{a(fk-jg)+e(jc-bk)+i(bg-cf)}\begin{bmatrix}d\\h\\l\end{bmatrix}\tilde{T}, \tag{19.157}$$

$$U=\frac{dfk-dgj+hcj-hbk+lbg-lcf}{a(fk-jg)+e(jc-bk)+i(bg-cf)}\tilde{T}=C_1\tilde{T}, \tag{19.158}$$

$$V=\frac{dgi-dek+hak-hci+lce-lag}{a(fk-jg)+e(jc-bk)+i(bg-cf)}\tilde{T}=C_2\tilde{T}, \tag{19.159}$$

$$W=\frac{dej-dif+hbi-haj+laf-lbe}{a(fk-jg)+e(jc-bk)+i(bg-cf)}\tilde{T}=C_3\tilde{T}, \tag{19.160}$$

$$\tilde{\varepsilon}_x=j2\pi\xi C_1\tilde{T},\ \tilde{\varepsilon}_y=j2\pi\eta C_2\tilde{T},\ \tilde{\varepsilon}_z=j2\pi\omega C_3\tilde{T}, \tag{19.161}$$

and shear components of strain-displacement are:

$$\tilde{\gamma}_{xy}=j2\pi(\eta C_1+\xi C_2)\tilde{T},\ \tilde{\gamma}_{yz}=j2\pi(\omega C_2+\eta C_3)\tilde{T}, \text{ and } \tilde{\gamma}_{zx}=j2\pi(\xi C_3+\omega C_1)\tilde{T}. \tag{19.162}$$

From matrix equation relating strain to stress is the same for both spatial and frequency domains. From Eqs. (19.51–19.53)

$$\tilde{\sigma}_x = \frac{E}{(1+\mu)(1-2\mu)}[(1-\mu)\tilde{\varepsilon}_x + \mu\tilde{\varepsilon}_y + \mu\tilde{\varepsilon}_z] - \frac{E\alpha\tilde{T}}{(1-2\mu)}, \tag{19.163}$$

$$\tilde{\sigma}_y = \frac{E}{(1+\mu)(1-2\mu)}[\mu\tilde{\varepsilon}_x + (1-\mu)\tilde{\varepsilon}_y + \mu\tilde{\varepsilon}_z] - \frac{E\alpha\tilde{T}}{(1-2\mu)}, \tag{19.164}$$

$$\tilde{\sigma}_z = \frac{E}{(1+\mu)(1-2\mu)}[\mu\tilde{\varepsilon}_x + \mu\tilde{\varepsilon}_y + (1-\mu)\varepsilon_z] - \frac{E\alpha\tilde{T}}{(1-2\mu)}, \tag{19.165}$$

$$\tilde{\sigma}_x(\xi, \eta, \omega) = -\frac{E\alpha(\eta^2+\omega^2)}{\xi^2+\eta^2+\omega^2}\tilde{T}(\xi, \eta, \omega) = H_x(\xi, \eta, \omega)\tilde{T}(\xi, \eta, \omega), \qquad \tilde{\sigma}_x(0, 0, 0) = 0, \tag{19.166}$$

$$\tilde{\sigma}_y(\xi, \eta, \omega) = -\frac{E\alpha(\omega^2+\xi^2)}{\xi^2+\eta^2+\omega^2}\tilde{T}(\xi, \eta, \omega) = H_y(\xi, \eta, \omega)\tilde{T}(\xi, \eta, \omega), \qquad \tilde{\sigma}_y(0, 0, 0) = 0, \tag{19.167}$$

$$\tilde{\sigma}_z(\xi, \eta, \omega) = -\frac{E\alpha(\xi^2+\eta^2)}{\xi^2+\eta^2+\omega^2}\tilde{T}(\xi, \eta, \omega) = H_z(\xi, \eta, \omega)\tilde{T}(\xi, \eta, \omega), \qquad \tilde{\sigma}_z(0, 0, 0) = 0, \tag{19.168}$$

$$\tilde{\tau}_{xy}(\xi, \eta, \omega) = \frac{E\alpha\xi\eta}{\xi^2+\eta^2+\omega^2}\tilde{T}(\xi, \eta, \omega) = H_{xy}(\xi, \eta, \omega)\tilde{T}(\xi, \eta, \omega), \qquad \tilde{\tau}_{xy}(0, 0, 0) = 0, \tag{19.169}$$

$$\tilde{\tau}_{yz}(\xi, \eta, \omega) = \frac{E\alpha\eta\omega}{\xi^2+\eta^2+\omega^2}\tilde{T}(\xi, \eta, \omega) = H_{yz}(\xi, \eta, \omega)\tilde{T}(\xi, \eta, \omega), \qquad \tilde{\tau}_{yz}(0, 0, 0) = 0, \tag{19.170}$$

$$\tilde{\tau}_{zx}(\xi, \eta, \omega) = \frac{E\alpha\omega\xi}{\xi^2+\eta^2+\omega^2}\tilde{T}(\xi, \eta, \omega) = H_{zx}(\xi, \eta, \omega)\tilde{T}(\xi, \eta, \omega), \qquad \tilde{\tau}_{zx}(0, 0, 0) = 0. \tag{19.171}$$

We note that all transfer functions are hermetian, satisfying the condition:

$$\mathrm{Re}[H(\xi, \eta, \omega)] = \mathrm{Re}[H(-\xi, -\eta, -\omega)] \text{ and } \mathrm{Im}[H(\xi, \eta, \omega)] = -\mathrm{Im}[H(-\xi, -\eta, -\omega)], \tag{19.172}$$

where $H(\xi, \eta, \omega)$ is the Fourier transform of a strictly real function $h(x, y, z)$. The Fourier transform of temperature, $\tilde{T}(\xi, \eta, \omega)$, is also hermetian as $T(x, y, z)$ is real. The product of two hermetian functions is hermetian, so all stress terms are hermetian as well.

19.4 Optical Effects Of Stress

The stress induces a local change in index that causes polarization-dependent. We shall assume that the unstressed material is optically isotropic such as NaCl or glass in the sense that the index is the same in all directions. In most cases of interest the magnitude of index change is relatively weak (as opposed to a crystal like calcite). The stress alters the permeability tensor by way of the 6×6 photoelastic tensor. Consider the tensor of perturbations to the electric permeability

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(0) + \begin{bmatrix} \Delta\varepsilon_1 & \Delta\varepsilon_6 & \Delta\varepsilon_5 \\ \Delta\varepsilon_6 & \Delta\varepsilon_2 & \Delta\varepsilon_4 \\ \Delta\varepsilon_5 & \Delta\varepsilon_4 & \Delta\varepsilon_3 \end{bmatrix}. \tag{19.173}$$

Where the symmetry results in six independent coefficients. We may arrange both the stress and electric permeability coefficients as a six-element column vector. For cubic (23, m3) material, the equation for permeability change as a function of stress is[Table 15, Ref. 3]

$$\begin{bmatrix} \Delta\varepsilon_1 \\ \Delta\varepsilon_2 \\ \Delta\varepsilon_3 \\ \Delta\varepsilon_4 \\ \Delta\varepsilon_5 \\ \Delta\varepsilon_6 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ c_{13} & c_{11} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{13} & c_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{44} \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{bmatrix}. \tag{19.174}$$

As compared with cubic material, the isotropic material has $c_{13} = c_{12}$ and $c_{44} = c_{11} - c_{12}$, so we can write the equations for cubic material and apply the coefficient simplifications for isotropic material.

Given $\varepsilon = 1/n^2$ we have $\Delta\varepsilon = -2\Delta n/n_0^3$ so we have the matrix equation

$$\Delta \boldsymbol{n} = -\frac{1}{2} n_0^3 \boldsymbol{\tau\sigma}. \tag{19.175}$$

In general in anisotropic media the electric field is not orthogonal to the phase propagation vector $\boldsymbol{k}$ (the normal to the wavefront). However for the small perturbations typical of stress birefringence imposed on an isotropic medium, the rotation is quite small. We may estimate the rotation by

$$\mathbf{R}_{\Delta\boldsymbol{\varepsilon}} = \frac{\varepsilon \mathrm{I} + \Delta\boldsymbol{\varepsilon}}{\det|\varepsilon \mathrm{I} + \Delta\boldsymbol{\varepsilon}|}. \tag{19.176}$$

For perturbations on the order of 10^{-4} or less (typically the stress-induced perturbations are quite small) the rotation of the electric vector $\boldsymbol{E}$ relative to $\boldsymbol{k}$ is essentially negligible and may be neglected. This also implies that the walk-off is negligible. Under these assumptions, we may calculate the effect of the index perturbation tensor on any arbitrary state of polarization designated by the vector $\boldsymbol{a}(0)$ by rotating into the

principle coordinate system of $\Delta \boldsymbol{n}$. We may use a method such as the Jacobi transformation to decompose $\Delta \boldsymbol{n}$ into diagonal form:

$$\Delta \boldsymbol{n} = \mathbf{R}^{-1} \begin{bmatrix} \Delta n_1 & 0 & 0 \\ 0 & \Delta n_2 & 0 \\ 0 & 0 & \Delta n_3 \end{bmatrix} \mathbf{R} . \tag{19.177}$$

Given propagation along the z-axis the effect of stress birefringence is

$$\boldsymbol{e}(z) = \mathbf{R}^{-1} \begin{bmatrix} e^{j\frac{2\pi}{\lambda} n_1 z} & 0 & 0 \\ 0 & e^{j\frac{2\pi}{\lambda} n_2 z} & 0 \\ 0 & 0 & e^{j\frac{2\pi}{\lambda} n_3 z} \end{bmatrix} \mathbf{R} \boldsymbol{e}(0) \tag{19.178}$$

and stress birefringence may be considered as a combination of optical path difference and shearing (walk-off) aberrations. At each local point the general index may be defined in terms of the index ellipsoid

$$\frac{x'^2}{n_x^2} + \frac{y'^2}{n_y^2} + \frac{z'^2}{n_z^2} = 1 , \tag{19.179}$$

where (x', y', z') are in an arbitrarily rotated coordinate system, defined by three rotation angles, so that the state of index is defined by six coefficients.

For an arbitrary propagation direction, there are two eigenstates of polarization with, in general, two distinct indices of refraction corresponding to eigenvalues. In the case that we know the two eigenstates of polarization $\boldsymbol{e}_1$ $\boldsymbol{e}_2$ and associated indices of refraction n_1 and n_2 , then we may write

$$\boldsymbol{e}_1(z) = \boldsymbol{e}_1(0) e^{j\frac{2\pi}{\lambda} n_1 z} , \boldsymbol{e}_2(z) = \boldsymbol{e}_2(0) e^{j\frac{2\pi}{\lambda} n_2 z} . \tag{19.180}$$

where

$$\boldsymbol{e} = \begin{bmatrix} e_x \\ e_y \\ e_z \end{bmatrix} . \tag{19.181}$$

We can write the matrix equation for the transformation of the two polarization states as

$$\left[\boldsymbol{e}_1(z)\ \boldsymbol{e}_2(z)\right] = \begin{bmatrix} e^{j\frac{2\pi}{\lambda}n_1 z} & 0 & 0 \\ 0 & e^{j\frac{2\pi}{\lambda}n_2 z} & 0 \\ 0 & 0 & e^{j\frac{2\pi}{\lambda}n_3 z} \end{bmatrix} \left[\boldsymbol{e}_1(0)\ \boldsymbol{e}_2(0)\right]. \tag{19.182}$$

Any two orthogonal vectors may be propagated by rotating them into the principle coordinate system for application of the media effects and then rotating back out:

$$\left[\boldsymbol{a}_1(z)\ \boldsymbol{a}_2(z)\right] = \mathbf{R}^{-1} \begin{bmatrix} e^{j\frac{2\pi}{\lambda}n_1 z} & 0 & 0 \\ 0 & e^{j\frac{2\pi}{\lambda}n_2 z} & 0 \\ 0 & 0 & e^{j\frac{2\pi}{\lambda}n_3 z} \end{bmatrix} \mathbf{R} \left[\boldsymbol{a}_1(0)\ \boldsymbol{a}_2(0)\right]. \tag{19.183}$$

We may express the media matrix as a matrix exponential and evaluate the matrix exponential as a Taylor series

$$e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}}{2!} + \frac{\mathbf{A}}{3!} + \ldots. \tag{19.184}$$

Exponentiation of a diagonal matrix is a special case, so we have the relationship:

$$e^{j\frac{2\pi z}{\lambda}\begin{bmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{bmatrix}} = \begin{bmatrix} e^{j\frac{2\pi}{\lambda}n_1 z} & 0 & 0 \\ 0 & e^{j\frac{2\pi}{\lambda}n_2 z} & 0 \\ 0 & 0 & e^{j\frac{2\pi}{\lambda}n_3 z} \end{bmatrix}. \tag{19.185}$$

The expression for propagation of any arbitrary polarization state $\boldsymbol{a}$ is

$$\boldsymbol{a}(z) = \mathbf{R}^{-1} e^{j\frac{2\pi z}{\lambda}\begin{bmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{bmatrix}} \mathbf{R}\boldsymbol{a}(0) = \mathbf{R}^{-1} e^{j\frac{2\pi z}{\lambda}\mathbf{N}_0} \mathbf{R}\boldsymbol{a}(0), \tag{19.186}$$

where $\mathbf{R}$ represents a rotation into the diagonalized coordinate system and the index matrix in the orthogonal coordinate system is

$$\mathbf{N}_0 = \begin{bmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{bmatrix}. \tag{19.187}$$

For a diagonal matrix $\mathbf{N}_0$

$$\mathbf{R}^{-1} e^{j\frac{2\pi z}{\lambda}\mathbf{N}_0}\mathbf{R} = e^{j\frac{2\pi z}{\lambda}\mathbf{R}^{-1}\mathbf{N}_0\mathbf{R}}. \tag{19.188}$$

So that the arbitrary vector $\boldsymbol{a}$ may be propagated by

$$\boldsymbol{a}(z) = e^{j\frac{2\pi z}{\lambda}\mathbf{R}^{-1}\mathbf{N}_0\mathbf{R}}\boldsymbol{a}(0), \tag{19.189}$$

where the rotation matrices are moved into the exponent. The expression for the general index matrix $\mathbf{N}$ is

$$\mathbf{N} = \mathbf{R}^{-1}\mathbf{N}_0\mathbf{R}. \tag{19.190}$$

As any symmetric matrix $\mathbf{N}$ may be represented in the diagonal form with rotation matrices as indicated in Eq. (19.190), we may write the propagation equation in terms of the general index matrix for any arbitrary polarization state $\boldsymbol{a}$:

$$\boldsymbol{a}(z) = e^{j\frac{2\pi z}{\lambda}\mathbf{N}}\boldsymbol{a}(0) = \boldsymbol{a}(0)e^{j\frac{2\pi z}{\lambda}\mathbf{N}}, \tag{19.191}$$

noting that for a symmetric matrix the order of multiplication is unimportant. We see from Eq. (19.191) that it is not strictly necessary to solve for the eigenvectors in order to propagate through the anisotropic medium.

In the particular case of weak birefringence, we may write the index matrix in terms of perturbation values:

$$\mathbf{N} = n_0\mathbf{I} + \boldsymbol{\Delta}, \text{ where } \boldsymbol{\Delta} = \begin{bmatrix} \Delta_1 & \Delta_6 & \Delta_5 \\ \Delta_6 & \Delta_2 & \Delta_4 \\ \Delta_5 & \Delta_4 & \Delta_3 \end{bmatrix}. \tag{19.192}$$

The matrix exponential of Eq. (19.191) may be written in terms of an ordinary scalar exponential of n_0 and an exponential of the perturbation matrix $\boldsymbol{\Delta}$ as defined in Eq. (19.192). Combining Eqs. (19.191) and (19.192),

$$\boldsymbol{a}(z) = \boldsymbol{a}(0)e^{j\frac{2\pi z}{\lambda}n_0}e^{j\frac{2\pi z}{\lambda}\boldsymbol{\Delta}}. \tag{19.193}$$

For light propagating primarily along the z-axis, we may simplify the 3×3 matrix $\mathbf{\Delta}$ by taking advantage of the fact that

$$\boldsymbol{a} = \begin{bmatrix} a_x \\ a_y \\ 0 \end{bmatrix} \tag{19.194}$$

and the anisotropy is weak. For these conditions we need consider only the upper 2×2 submatrix of $\mathbf{N}$ to form $\boldsymbol{N}_2$, as the z-component remains negligible for weak index perturbations,

$$\mathbf{N}_2 = n_0 \mathbf{I} + \begin{bmatrix} \Delta_1 & \Delta_6 \\ \Delta_6 & \Delta_2 \end{bmatrix} = \begin{bmatrix} \Delta_x & \Delta_{xy} \\ \Delta_{xy} & \Delta_y \end{bmatrix}, \tag{19.195}$$

where we have indicated a notation in x-y-z. We can identify the 2×2 perturbation submatrix as:

$$\mathbf{\Delta}_2 = \begin{bmatrix} \Delta_x & \Delta_{xy} \\ \Delta_{xy} & \Delta_y \end{bmatrix}. \tag{19.196}$$

We only need to solve for the three coefficients Δ_x, Δ_y, and Δ_{xy}. The polarization transformation may be written as a Jones calculus operator:

$$\mathbf{J} = e^{j\frac{2\pi z}{\lambda}\mathbf{\Delta}_2}. \tag{19.197}$$

If the index perturbations are weak and/or the propagation length is short, the matrix exponential may be evaluated using only the first order term of the Taylor series:

$$e^{j\frac{2\pi z}{\lambda}\mathbf{\Delta}_2} \approx \mathbf{I} + j\frac{2\pi z}{\lambda}\mathbf{\Delta}_2, \tag{19.198}$$

$$\boldsymbol{a}(z) \approx e^{j\frac{2\pi z}{\lambda}n_0}\left(\mathbf{I} + j\frac{2\pi z}{\lambda}\mathbf{\Delta}_2\right)\boldsymbol{a}(0)\ . \tag{19.199}$$

In this case, the Jones calculus operator for index perturbation is

$$\mathbf{J} \approx \mathbf{I} + j\frac{2\pi z}{\lambda}\mathbf{\Delta}_2. \tag{19.200}$$

19.4.1 Photoelastic Tensor

The index perturbation matrix may be calculated from the stress tensor by the photoelastic tensor, a 6×6 matrix. For a material that is isotropic, only two independent photoelastic coefficients are needed for the isotropic material. The photoelastic tensor for cubic material is

$$\begin{bmatrix} \Delta_x \\ \Delta_y \\ \Delta_z \\ \Delta_{yz} \\ \Delta_{zx} \\ \Delta_{xy} \end{bmatrix} = -\frac{n_0^3}{2} \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \\ \varepsilon_{yz} \\ \varepsilon_{zx} \\ \varepsilon_{xy} \end{bmatrix} = -\frac{n_0^3}{2} \begin{bmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ c_{13} & c_{11} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{13} & c_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{44} \end{bmatrix} \begin{bmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{yz} \\ \tau_{zx} \\ \tau_{xy} \end{bmatrix}. \tag{19.201}$$

For isotropic material only two coefficients are needed

$$\begin{bmatrix} \Delta_x \\ \Delta_y \\ \Delta_z \\ \Delta_{yz} \\ \Delta_{zx} \\ \Delta_{xy} \end{bmatrix} = -\frac{n_0^3}{2} \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \\ \varepsilon_{yz} \\ \varepsilon_{zx} \\ \varepsilon_{xy} \end{bmatrix} = -\frac{n_0^3}{2} \begin{bmatrix} c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{12} & c_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{11}-c_{12} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{11}-c_{12} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{11}-c_{12} \end{bmatrix} \begin{bmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{yz} \\ \tau_{zx} \\ \tau_{xy} \end{bmatrix}. \tag{19.202}$$

As compared with cubic material, the isotropic material has $c_{13} = c_{12}$ and $c_{44} = c_{11} - c_{12}$, so we can write the equations for cubic material and apply the coefficient simplifications for isotropic material. Since we only need Δ_x, Δ_y, and Δ_{xy}, we may simplify Eq. (19.202):

$$\begin{bmatrix} \Delta_x \\ \Delta_y \\ \Delta_{xy} \end{bmatrix} = -\frac{n_0^3}{2} \begin{bmatrix} c_{11} & c_{12} & c_{13} & 0 \\ c_{13} & c_{11} & c_{12} & 0 \\ 0 & 0 & 0 & c_{44} \end{bmatrix} \begin{bmatrix} \sigma_x \\ \sigma_y \\ \sigma_{xy} \\ \tau xy \end{bmatrix}. \tag{19.203}$$

Equation Eq. (19.203) is a linear equation, consisting of coefficients applied to the four stress values. We may apply the same equation in frequency space for cubic crystals

$$\begin{bmatrix} \tilde{\Delta}_x \\ \tilde{\Delta}_y \\ \tilde{\Delta}_{xy} \end{bmatrix} = -\frac{n_0^3}{2} \begin{bmatrix} c_{11} & c_{12} & c_{13} & 0 \\ c_{13} & c_{11} & c_{12} & 0 \\ 0 & 0 & 0 & c_{44} \end{bmatrix} \begin{bmatrix} \tilde{\sigma}_x \\ \tilde{\sigma}_y \\ \tilde{\sigma}_{xy} \\ \tilde{\tau}_{xy} \end{bmatrix}. \tag{19.204}$$

For isotropic material

$$\begin{bmatrix} \tilde{\Delta}_x \\ \tilde{\Delta}_y \\ \tilde{\Delta}_{xy} \end{bmatrix} = -\frac{n_0^3}{2} \begin{bmatrix} c_{11} & c_{12} & c_{12} & 0 \\ c_{12} & c_{11} & c_{12} & 0 \\ 0 & 0 & 0 & c_{11} - c_{12} \end{bmatrix} \begin{bmatrix} \tilde{\sigma}_x \\ \tilde{\sigma}_y \\ \tilde{\sigma}_{xy} \\ \tilde{\tau}_{xy} \end{bmatrix}. \tag{19.205}$$

A linear equation of hermetian functions is also hermetian, so $\tilde{\Delta}_x$, $\tilde{\Delta}_y$, and $\tilde{\Delta}_{xy}$ are hermetian and Δ_x, Δ_y, and Δ_{xy} are guaranteed to be real. The resulting inverse Fourier transform yields the index matrix for all points in the volume:

$$\boldsymbol{\Delta}(x, y, z) = \begin{bmatrix} \Delta_x(x, y, z) & \Delta_{xy}(x, y, z) \\ \Delta_{xy}(x, y, z) & \Delta_z(x, y, z) \end{bmatrix}. \tag{19.206}$$

The effect of the index variation is integrated over the volume along the z-direction:

$$\boldsymbol{a}(z) = \boldsymbol{a}(0) e^{j\frac{2\pi z}{\lambda} n_0 L} e^{j\frac{2\pi}{\lambda}\int_0^L \boldsymbol{\Delta}(x, y, x)\, dz}. \tag{19.207}$$

19.4.2 Simplifying the 3D Calculations

Equation (19.207) indicates that the optical effect of the birefringence is due to the integration of the index matrix along the z-axis.

$$\boldsymbol{\Delta}_L(x, y) = \int_0^L \boldsymbol{\Delta}(x, y, x)\, dz. \tag{19.208}$$

Hence we do not actually need the definition of index matrix over the volume but only the z-integrated value. The integrated value may be calculated much more easily as only 2D calculations are required.

Let the z-integrated temperature

$$T_L(x, y) = \int_0^L T(x, y, z)\, dz, \tag{19.209}$$

in units of degree-centimeters. We form the two-dimensional Fourier transform $\tilde{T}_L(\xi, \eta)$ by setting $\omega = 0$,

$$\tilde{\sigma}_x(\xi, \eta) = -\frac{E\alpha\eta^2}{\xi^2 + \eta^2} \tilde{T}_L(\xi, \eta) = H_x(\xi, \eta) \tilde{T}_L(\xi, \eta), \qquad \tilde{\sigma}_x(0, 0) = 0, \tag{19.210}$$

$$\tilde{\sigma}_y(\xi,\eta) = -\frac{E\alpha\xi^2}{\xi^2+\eta^2}\tilde{T}_L(\xi,\eta) = H_y(\xi,\eta)\tilde{T}_L(\xi,\eta), \qquad \tilde{\sigma}_y(0,0) = 0, \tag{19.211}$$

$$\tilde{\tau}_{xy}(\xi,\eta) = \frac{E\alpha\xi\eta}{\xi^2+\eta^2}\tilde{T}_L(\xi,\eta) = H_{xy}(\xi,\eta)\tilde{T}_L(\xi,\eta), \qquad \tilde{\tau}_{xy}(0,0) = 0, \tag{19.212}$$

$$\tilde{\tau}_{yz}(\xi,\eta) = 0, \tag{19.213}$$

$$\tilde{\tau}_{zx}(\xi,\eta) = 0. \tag{19.214}$$

The stress-induced index coefficients are found with the reduced photoelastic tensor for cubic material as before and applied to the fourier transform of the z-integrated temperature distribution

$$\begin{bmatrix} \tilde{\Delta}_{x_L} \\ \tilde{\Delta}_{y_L} \\ \tilde{\Delta}_{xy_L} \end{bmatrix} = -\frac{n_0^3}{2}\begin{bmatrix} c_{11} & c_{12} & 0 \\ c_{13} & c_{11} & 0 \\ 0 & 0 & c_{44} \end{bmatrix}\begin{bmatrix} H_x \\ H_y \\ H_{xy} \end{bmatrix}\tilde{T}_L. \tag{19.215}$$

Equation (19.215) gives the two-dimensional index perturbation matrix yielding the OPD matrix $\boldsymbol{\Delta}_L$. Expanding the matrix calculations into the three equations:

$$\tilde{\Delta}_{x_L} = \frac{E\alpha n_0^3}{2(\xi^2+\eta^2)}(c_{12}\xi^2 + c_{11}\eta^2)\tilde{T}_L, \tag{19.216}$$

$$\tilde{\Delta}_{y_L} = \frac{E\alpha n_0^3}{2(\xi^2+\eta^2)}(c_{13}\eta^2 + c_{11}\xi^2)\tilde{T}_L, \tag{19.217}$$

$$\tilde{\Delta}_{xy_L} = -\frac{E\alpha n_0^3}{2(\xi^2+\eta^2)}c_{44}\xi\eta\tilde{T}_L. \tag{19.218}$$

Photoelastic coefficients for some cubic materials are given in Table 19.1.

Table. 19.1. Photoelastic coefficients for some cubic materials (units = 10^{-12} m^2/newton). From J. F. Nye, *Physical Properties of Crystals*[3].

Crystal	Class	c_{11}	c_{12}	c_{13}	c_{44}
sodium chloride	m3m	0.25	1.46	1.46	-0.85
diamond	m3m	-0.43	0.37	0.37	-0.27

References

1. Pei Chi Chou and Nicholas J. Pagano, *Elasticity, Tensor, Dyadic and Engineering Approaches*, Dover Publications, Inc. (1967).
2. Zhilun Xu, *Applied Elasticity*, John Wiley & Sons, (1992).
3. J. F. Nye, *Physical Properties of Crystals*, Oxford Science publications (1985).
4. Amnon Yariv and Pochi Yeh, *Optical Waves in Crystals*, John Wiley & Sons (1984).

20. Integration of Geometrical and Physical Optics

This chapter outlines a method of combining geometrical and physical optics. We shall be primarily concerned with well-behaved, i.e., imaging optical systems. Such systems image a region in object space into image space. Generally the best imaging is found for a limited range of object conjugate (the object to lens distance) and form a limited range of object angles. The imaging is never perfect, suffering from truncation of the beam by the apertures and aberrations of the optical system.

The imaging property of the well-behaved system defines the approach for modeling well-behaved optical systems from their geometrical optics behavior. If the imaging were perfect, then the input beam would be geometrically modified into an output beam in image space. Considered between the conjugate planes, the ideal system simply applies magnification, change of direction, and change of phase radius of curvature — no diffraction effects, no aberration, and not aperture clipping. The real optical system, necessarily imperfect is a perturbation from the ideal system.

The representation of the physical optics beam is somewhat more complex than is generally used in geometrical optic calculations. We can, however, establish how all aspects of the physical optics are transformed by the optical system. In the GLAD program the optical beam may be represented by

$$a_i(x, y)\,,\ b_j(x, y)\,,\ \boldsymbol{r},\ \mathrm{K},\ R_x\,,\ R_y\,,\ \omega, \tag{20.1}$$

$a_i(x, y)$ complex amplitude distribution of in the $\hat{\boldsymbol{i}}$-direction,

$b_j(x, y)$ complex amplitude distribution of in the $\hat{\boldsymbol{j}}$-direction,

$\boldsymbol{r}$ location of the center of the beam in global coordinates,

$\mathbf{K}$ coordinate system for the beam directions $\mathbf{K} = \left[\hat{\boldsymbol{i}}\ \hat{\boldsymbol{j}}\ \hat{\boldsymbol{k}}\right]$,

R_x, R_y Radii of reference surface,

ω frequency, establishes wavelength.

Figure 20.1 illustrates schematically the transformations due to imaging optical system. The changes are summarized as

aberration $a_i(x, y) \Rightarrow e^{iW(x, y)} a_i(x, y)\,,$

aberration $a_j(x, y) \Rightarrow e^{iW(x, y)} a_j(x, y)\,,$

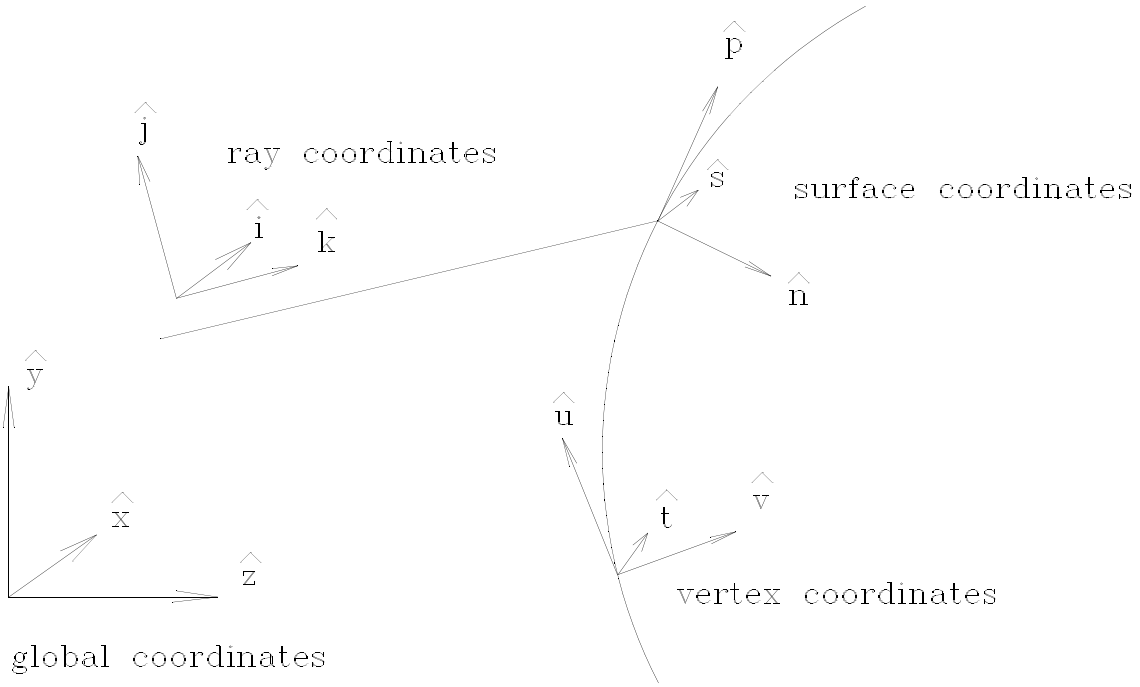


Fig. 20.1. The global, ray, vertex, and surface coordinate system.

coordinate shift	$\boldsymbol{r} \Rightarrow \boldsymbol{r} + \Delta\boldsymbol{r}$,
rotation and parity change	$\mathbf{K} \Rightarrow \mathbf{RK}$,
phase radius change	$R \Rightarrow \dfrac{A + \dfrac{B}{R}}{C + \dfrac{D}{R}}$,
aperture stop	diffraction to and from aperture stop.

where $W(x, y)$ is the phase aberration, $\Delta\boldsymbol{r}$ is the shift position between object and image, **R** represents the rotation of the ray coordinate system (and parity change for odd-numbered mirror systems), and ABCD is the paraxial system in the x- or y-direction.

In many lens designs there is a well defined aperture stop (at least for the on-axis optical beam), unlike general physical optics systems where clipping apertures may be more arbitrarily located. In addition to the aperture stop, an off-axis beam may be clipped by apertures ahead of or following the aperture stop. Figure 20.2 illustrates a representative system. Generally, there is a well-defined region where the beam is expanded and the elements are relatively close, such that the elements are effectively in the near-field with respect to each other. This is an important condition for geometrical design to be valid. Apertures at different axial points in the expanded region are effectively co-located in terms of diffraction calculations. The

various apertures in the expanded region may be collapsed into the entrance pupil. The aperture stop may be accurately included by using diffraction propagation from the object to the entrance pupil, applying the entrance pupil aperture as the aperture stop and including phase aberrations, and propagating back to the object point. The object point may be transformed to its conjugate in image space by a simple ABCD matrix of the form:

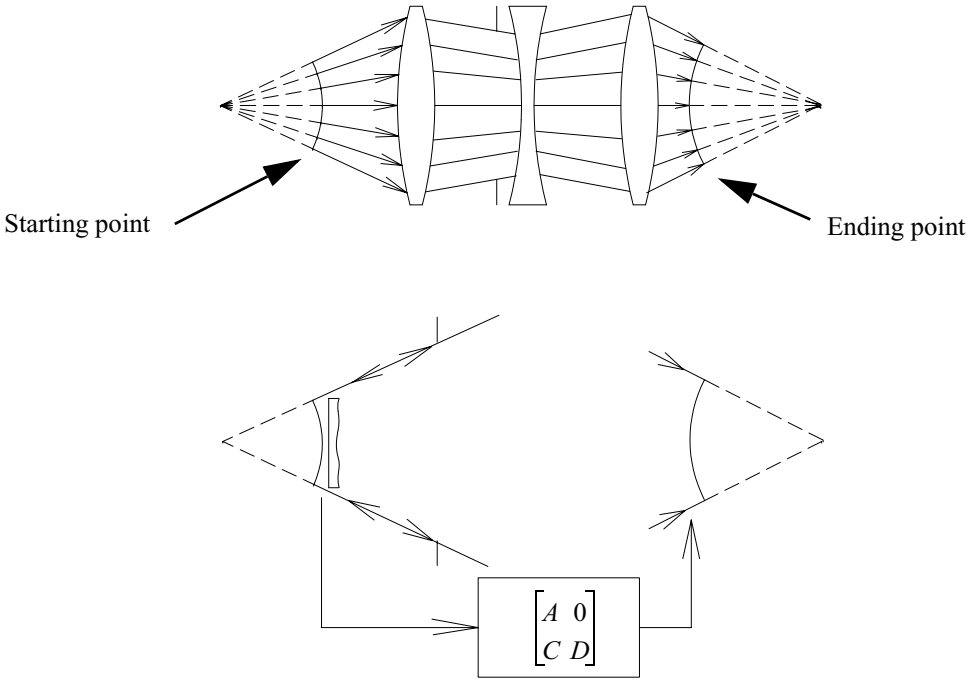


Fig. 20.2. The aberrations of the optical system are determined by probing with rays. Generally, the rays will be started at some intermediate point on the beam in object space and terminating in image space. The system may be represented by an aberration plate and the paraxial optics behavior.

$$\text{ABCD} = \begin{bmatrix} A & 0 \\ C & D \end{bmatrix}, \tag{20.2}$$

where $A = M$ (M is the magnification), $D = 1/M$, and C represents the optical power. $B = 0$ indicates there is no diffraction propagation in going between conjugate points.

In the general cae, tilts and decenters are allowed and the aberrations are calculated correspondingly. See Fig. 20.3.

The aberrations may be determined by probing the system with rays. The rays are constructed to be normal to the reference surface. Ideally the rays are constructed to be normal to the wavefront but in order to probe the system with rays it is generally not necessary to make a distinction between the wavefront normals and the reference surface normals.

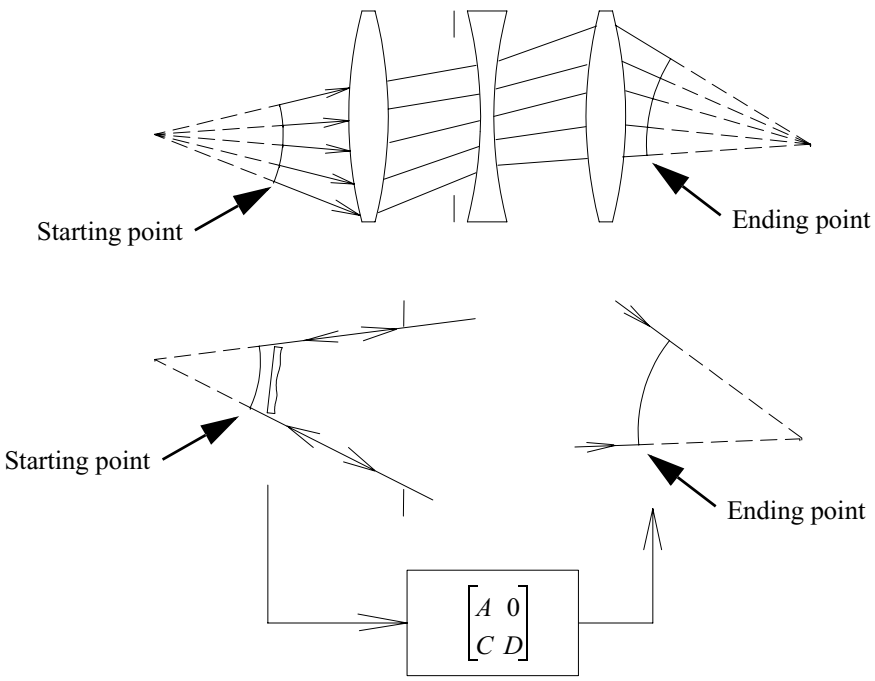


Fig. 20.3. tIn the general case, the equivalent optical system must include global coordinate shifts and rotations to take into account of the displacement of the starting point to the ending point.

Index

A

ABCD
  matrix method 94
ABCD matrix method 63, 68, 70, 170, 172
aberration
  aberration 80
  atmospheric 215
  Kolmogorov 81, 215
  random 80–81
  random aberration 78
  Zernike 77
aberrations 345
adaptive optic
  actuators 222
adaptive optics 215, 219
  actuators 224–225
  influence function 221, 224, 226
  influence functions 222
  printthrough 225
Airy pattern 51
aliasing 27, 43, 48–49, 51, 225
analysis, end-to-end system 15
anamorphic 172
annular beams 292
apertures 170
atmospheric
  index of refraction 230
  inner and outer scale 216
atomospheric
  refractive index structure 216
autocorrelation
  width 80, 216
axicon 285
axicons
  afocal mode 292
  hot tip 289
  polarization 292
  reflaxicons 299
  resonator 297
  waxicons 299

B

beam propagation method (BPM) 69–70
Bessel function 36
Bragg condition 264–265
Brillouin 18

C

chief ray 179
chirality 192
coherent transfer function 307
comb function 44–46, 224–225
components 169
constitutive matrix 316
coordinate system
  global 176
  ray 176
  surface 176
  vertex 176

D

downlink 215

E

eigenfunctions 24
eigenmode 94, 101
eigenvalue 97
eigenvalues 24
Euler angles 177

F

Fermat's Principle 189
FFT 15, 22, 43, 219, 286, 295
Fouier transform 46
Fourier transform 37, 225
Fresnel 249–250
  approximation 43
  diffraction 20
  number 30, 33, 51
  zones 34
Fried's parameter 81, 216–217

G

gain 111

Beer's Law 111
Franz-Nodvik 112
rate equation 113
ruby 120
three-level 120
gaussian beam 38–39
geometrical
optics 13
ray tracing 13, 179, 187
geometrical optics 345
GLAD 9, 15, 18, 20, 169, 175, 179, 182, 229, 239, 242, 248, 287, 313, 345
manuals 9
global coordinates 175, 177, 345
Goodman, D. S. 306
Goos-Hanchen effect 258
gradient index (GRIN) 67, 170
Grahm-Schmidt 221
Green's Function 329, 332
guardband 49

H

heat conduction 309–310
hermal blooming 234
Hermite gaussian 15, 38, 103
Hooke's Law 324–325
Hopkins, H. H. 306

I

impulse response function 27
index ellipsoid 336
induced dipole 265, 268

J

Jackson 265, 274
Jacobian matrix 278
Jones matrix 240

L

Lagrange multipliers 281–282
Laguerre-gaussian 103–104
least squares 220, 226
lenses 169

M

Marchand 68
Maxwell's equations 19
Maxwell's fisheye lens 68, 71
merit function 283
mirrors 169, 189
matrices 189, 191–192
mode coupling 270

N

nonlinear optics
four-wave mixing 133
Raman 132
second harmonic generation 156, 158
transient Raman 159
Nyquist 45, 47

O

obscurations 170
optical parametric oscillator 138
optical path length 29
optical power 314
optimization 277
constraints 280, 282
dampled least squares 280
Jacobian 277
least squares 278

P

parity 192
partial coherence 305
phase conjugation 18
phased array 26
photoelastic tensor 340
point spread function 27, 32
Poison's Ratio 315
polarization 17, 20, 175, 239–240
aberrations 245
birefringence 242
distributed 241
Faraday effect 243
Fresnel transmission and reflection 245, 248
Kerr effect 244
optical acitvity 243
Pauli spin matrices 246
Pockels effect 244
rotation 246
single layer thin film 248
prisms 196
propagation
angular spectrum 24
atmospheric 215

coherent 306
control 51
diffraction 24, 29
far-field 36, 47
fast-Hankel 286
fast-Hankel transform 60, 295–296
finite difference 15, 59
free space 19
geometrical 25
near-field 37, 47, 49
radial 286
transfer function 26

Q

quadratic phase factor 27, 58

R

Raman 18
amplifier 15
Rayleigh distance 39
reflection 184
reflection coefficients 250
resonators 14, 91
bare-cavity 92
convergence 93
coupled 26, 175
g-parameters 98
stable 98–99, 102
unstable 106
rotation matrix 177
rotations 177

S

sampling 43, 47, 51, 312
second harmonic generation 18
Siegman 14, 63–64, 68–69, 93
smoothing 80, 225
space relay system 215
spatial filter 14, 215
split step 21
split step method 70–71
strain 317
energy 317
stress 309, 315
birefringence 315
stress/strain 316
sum-frequency generation (SFG) 149
surrogate gaussian beam 53, 56–58, 102

T

Talbot 35
imaging 29, 36
thermal blooming 229, 233
thermal coefficient of expansion 314
thermal conductivity 309–310
thermal modeling 309
thermally induced birefringence stress 315
thermo-optical effects 313
three dimensional 17, 20
transmission coefficients 249

U

uplink 215

W

wave equation 19
waveguide grating coupler 261
Welford 273–274

Y

Yariv 63, 267, 274