

**PROPAGATION OF LIGHT
IN
TAPERED GRADED-INDEX MEDIA**

Derek Carmelo Bertilone

**A thesis submitted for the degree of
Doctor of Philosophy
at the Australian National University
Canberra
April 1988**

Declaration

This thesis is an account of research undertaken in the Optical Sciences Centre at the Research School of Physical Sciences, Australian National University, between April 1984 and April 1988, for the degree of Doctor of Philosophy.

The material presented in chapters 2 and 8 is the result of research undertaken in collaboration with Professor C. Pask.

The material in chapter 3 is the result of research undertaken in collaboration with Dr A. Ankiewicz and Professor C. Pask.

The material in chapter 7 is the result of research undertaken in collaboration with Dr J.D. Love and Professor C. Pask.

None of the work presented here has been submitted to any other institution of learning for any degree.

Derek C. Bertilone

(Derek C. Bertilone)

Acknowledgements

My sincerest gratitude goes to my supervisor, Colin Pask. His enthusiasm and love for physics is evident in his wide interests and the sense of fun in which he approaches scientific problems. In addition, he is one of the finest teachers I have known, and his patience and friendliness is much appreciated.

Thanks also to John Love and Adrian Ankiewicz, who both contribute greatly to the friendly atmosphere of the Optical Sciences Centre. They are always ready to discuss new ideas and give freely of their own expert knowledge of optical fiber theory.

I also thank Andrew Stevenson, a good friend and fellow student, whose ideas know no bounds.

Thanks also to Diana Wallace for the excellent job in typing and editing this thesis.

For the past two years I have been fortunate to receive financial support by way of a cadetship with the Surveillance Research Laboratory, of the Defence Science and Technology Organization. I am very grateful for their generosity and support.

To my parents, words cannot express what I feel for their unending love and support.

Finally, I thank Donna, whose love and care at all times has not only kept me sane, but happy.

List of Publications

1. D. Bertilone, "Light Distribution in the Focal Regions of a Lens in a Parabolic Index Medium," Appl. Opt. 26, 306-317 (1987).
2. D. Bertilone and C. Pask, "Exact Ray Paths in a Graded-Index Taper," Appl. Opt. 26, 1189-1194 (1987).
3. D. Bertilone, A. Ankiewicz and C. Pask, "Wave Propagation in a Graded-Index Taper," Appl. Opt. 26, 2213-2221 (1987).
4. D. Bertilone, "Ray Propagation and Compression in a Strictly Adiabatic Taper," Opt. Quantum Electron. 19, 361-375 (1987).
5. D. Bertilone, "Modification of the Focused Field in a Parabolic Index Medium by Slow Axial Tapering," Appl. Opt. (in press).
6. D. Bertilone, J.D. Love and C. Pask, "Splicing of Optical Waveguides with Lossless GRIN Tapers," Opt. Quantum Electron. (submitted).
7. C. Pask and D. Bertilone, "New Model for Analysing Cone Elements in Visual Systems," Nature (submitted).

Abstract

Better understanding of the theory of light propagation in tapering graded-index structures is required for the design of novel optical devices. This thesis looks at some aspects of the behavior of light in media with refractive index graded both along and transverse to the direction of propagation. The emphasis is on guided-wave propagation, diffraction and self-focusing, with applications in optical electronics and vision.

Exact ray analysis and wave theory is given for several model tapers (including tapering quadratic media), and guided propagation is investigated in detail. The models elucidate many important principles underlying the physical behavior of light in tapering structures. An interesting feature of the solutions is the clarity with which they display the compression and reflection of light in the taper.

To investigate the effects of tapering on diffraction and self-focusing, we further develop the scalar theory of diffraction in quadratic media and present a detailed study of the 3D light distribution at the foci of a self-focused field. Extending the theory to slowly (but otherwise arbitrarily) tapered media, we show how the structure of the focused field is modified as a result of the tapering.

Some applications of the theory are presented. The low-loss splicing of dissimilar optical waveguides using strictly adiabatic tapers is analysed. Loss arises at the waveguide/taper junctions due to the different wavefront curvatures of the guided fields, and is quantified for the case of slab waveguides.

We suggest a model for the crystalline cone of the butterfly compound eye, which explains recently reported measurements. The model involves an optical principle that is new to vision: radiation-free guidance in GRIN tapers.

Notes on text

Within each chapter, equations and figures are numbered in order of appearance; e.g. Eqn. (7) or Fig. (3). To refer to an equation in a different chapter, we append the chapter number; e.g. Eqn. (3.7).

References are collected at the end of each chapter, and are labelled by superscript within the text.

TABLE OF CONTENTS

1. INTRODUCTION: GRADED-INDEX OPTICS AND TAPERED DIELECTRIC STRUCTURES

1.1	What are graded-index media?	1
1.2	Recent history of GRIN optics.	4
1.3	Applications of GRIN - now and future.	8
1.4	Tapered GRIN media.	10
1.4.1	Why taper GRIN rods?	11
1.4.2	Recent progress with theory.	13
1.5	Background theory: light propagation in graded-index media.	15
1.6	About this thesis.	19
1.7	References.	20

2. PROPAGATION IN A TAPERED GRADED-INDEX MEDIUM I: EXACT RAY ANALYSIS

2.1	Introduction.	31
2.2	The graded-index model.	32
2.3	The formalism.	34
2.3.1	The ray equations.	34
2.3.2	Initial conditions.	36
2.4	Ray path: R variable.	37
2.5	Ray path: θ variable.	39
2.6	Ray path: ϕ variable.	40
2.7	General features of the ray paths.	40
2.8	Ray solution of the GRIN taper.	42
2.8.1	Meridional rays.	44
2.9	Behavior of rays in the taper.	45
2.10	Conclusion.	48
2.11	References.	49

3. PROPAGATION IN A TAPERED GRADED-INDEX MEDIUM II: WAVE THEORY

3.1	Introduction.	50
3.2	Model of 2D taper.	51

3.3	Exact solution of the Helmholtz equation.	53
3.3.1	Small-Angle approximation.	55
3.3.2	Comparison with untapered medium.	56
3.4	Properties of the basis fields.	58
3.4.1	Comparison with geometrical optics.	59
3.4.2	Basis field as a transverse resonance.	62
3.5	Examples of applications of basis fields.	64
3.5.1	Fields launched down the complete taper.	64
3.5.2	Fields launched up the complete taper.	68
3.5.3	Fields launched in a section of taper.	69
3.6	Propagation inside a taper of circular geometry.	70
3.6.1	Exact solution of the Helmholtz equation.	71
3.6.2	The small-angle approximation.	72
3.7	Conclusion.	73
3.8	Appendix: limiting behavior of basis fields.	74
3.9	References.	77
4.	RAY PROPAGATION AND COMPRESSION IN A STRICTLY ADIABATIC TAPER	
4.1	Introduction.	78
4.2	Taper model.	79
4.3	The ray equations.	80
4.4	Ray propagation.	83
4.4.1	Ray path: R variable.	83
4.4.2	Ray path: θ variable.	85
4.4.3	Changing the external refractive index.	88
4.5	Ray compression.	90
4.6	Comparison with modal fields.	94
4.6.1	Radiation modes and refracting rays.	96
4.6.2.	Guided modes and bound rays.	98
4.7	3D taper.	99
4.7.1	Ray equations.	100
4.7.2	The θ equation.	101
4.7.3	Ray path characteristics.	103
4.7.4	The ϕ equation.	105
4.8	Conclusion.	105
4.9	References.	106

5. DIFFRACTION AND SELF-FOCUSING IN GRADED-INDEX MEDIA I: FORMALISM

5.1	Introduction.	108
5.2	General formulation and the paraxial approximation.	110
5.3	Diffraction in a quadratic-index medium.	113
5.3.1	The paraxial field.	116
5.3.2	Homogeneous medium limit.	116
5.4	Plane waves normally incident with thin lens in aperture.	117
5.4.1	Intensity along the axis.	118
5.4.2	Intensity in the geometrical focal plane.	124
5.4.3	Symmetry of the focused field.	125
5.5	Validity of the paraxial approximation.	126
5.5.1	Restrictions on parameters.	126
5.5.2	Restrictions on Z.	131
5.6	Geometrical optics.	132
5.7	Oblique incidence.	134
5.8	Comparison of focusing in quadratic and homogeneous media.	137
5.9	Conclusion.	139
5.10.1	Appendix A: homogeneous medium limit.	140
5.10.2	Appendix B: off axis incidence.	141
5.11	References.	144

6. DIFFRACTION AND SELF-FOCUSING IN GRADED-INDEX MEDIA II: SLOW AXIAL TAPERING

6.1	Introduction.	147
6.2	The tapered quadratic medium.	148
6.3	Adiabatic mode formalism.	150
6.4	Scalar diffraction in a slowly tapered GRIN medium.	153
6.5	Self-focusing of plane waves.	155
6.5.1	Intensity along the optical axis.	156
6.5.2	Intensity in geometrical focal plane.	158
6.6	Image transmission.	159
6.7	Comparison with non-adiabatic theory.	160
6.8	Conclusion.	164

6.9.1	Appendix A: adiabatic modes in a parabolically tapered quadratic medium.	165
6.9.2	Appendix B: validity of the diffraction analysis.	167
6.10	References.	169
7.	AN APPLICATION IN OPTICAL ELECTRONICS: SPLICING OF OPTICAL WAVEGUIDES WITH LOSSLESS GRIN TAPERS	
7.1	Introduction.	172
7.2	The Marcatili taper and the dielectric slab.	173
7.3	Gaussian approximation for the fundamental mode.	180
7.3.1	Taper design: spot-size matching.	182
7.4	Coupling efficiency formula.	183
7.4.1	Preliminaries.	183
7.4.2	Mode excitation.	184
7.4.3	Gaussian approximation.	187
7.5	Analysing the efficiency.	188
7.6	Conclusion.	194
7.7	References.	195
8.	AN APPLICATION IN VISION: A NEW MODEL FOR ANALYSING CONE ELEMENTS IN VISUAL SYSTEMS	
8.1	Introduction.	197
8.2	The cone model.	198
8.3	Meridional rays in the cone.	200
8.3.1	Weak grading assumption.	202
8.4	Special properties of the cone.	203
8.5	Focusing power of cone slices.	206
8.6	Compound eye of the butterfly.	212
8.7	Fitting model to butterfly data.	215
8.7.1	Check on parameter values.	218
8.7.2	Ray trajectories in the cone model.	220
8.8	Conclusion.	222
8.9	References.	224

CHAPTER 1

INTRODUCTION:

GRADED-INDEX OPTICS AND TAPERED DIELECTRIC STRUCTURES

This chapter serves two purposes - firstly we introduce the topic of graded-index optics, and give a brief review in order to bring some historical perspective to the role of tapering GRIN media. Secondly, we introduce the basic mathematical development which is assumed throughout the remainder of this thesis.

1.1 WHAT ARE GRADED-INDEX MEDIA?

A graded-index medium is one with an inhomogeneous distribution of refractive index. Symbolically, we write

$$n = n(\underline{R}) \quad (1)$$

to denote the functional relationship between the refractive index, n , and a point in the medium specified by the position vector $\underline{R}$. Often, there is some special symmetry which reduces the number of coordinates required to describe the index distribution. For example, spherical graded-index media have $n = n(R)$ where R denotes radial distance from some origin. Cylindrical graded-index media have $n = n(r, \phi)$ where (r, z, ϕ) are cylindrical polar coordinates, with z the cylinder axis and ϕ the azimuthal angle. Graded-Index media are also commonly referred to as GRIN, Gradient Index or Distributed Index media. In the special case of the index varying quadratically with distance in the transverse direction, the medium is sometimes referred to as Lens-Like or Self-Focusing.

GRIN media occur commonly in nature, with the atmosphere perhaps the most familiar example. At any point within the atmosphere, the refractive index depends, primarily, on the temperature, pressure and moisture content¹. All three quantities generally vary from point to point, leading to a smoothly graded refractive index profile. Many unusual atmospheric phenomena, mirages¹ being the best known example, result from the bending of the rays of light in this gradient. Related phenomena, such as the distortion in the shape of the sun and moon when observed near the horizon, are well-known to meteorologists and astronomers².

Graded media figure prominently in nature's approach to the design of animal eyes. In this context, the index grading usually results from an inhomogeneous concentration of protein in the membranes of some components of the eye. The crystalline lens of the human eye, for example, is composed of ultra-thin layers of material of different refractive index, "somewhat like a transparent onion"³. GRIN components also play a crucial role in image formation in the compound eyes of many invertebrates. Compound eyes can be classified, broadly, into apposition and superposition types (see, for example, the review by Land⁴). In the case of apposition eyes, each unit of the compound eye (referred to as an ommatidium) is optically isolated from every other ommatidium by light absorbing pigments. Light from a distant object is sampled by each ommatidium and brought to a focus on the endface of its associated rhabdom (light-sensitive photoreceptor)⁵⁻⁶. (An unusual exception has recently been discovered by Nilsson et al.⁷ and is discussed in section 1.4.1.) In superposition eyes, the crystalline cone in each ommatidium is a graded-index structure designed such that the cornea and cone together behave like an afocal telescope (an exception is found in a small number of

invertebrates which appear to have reflecting superposition optics⁴). A homogeneous region, the clear-zone, separates each cornea-cone unit from the retina, and partial images from each ommatidium are pieced together (superposed) to form a single, large image on the common retina. Afocal optics is essential for this approach⁸, and the crucial role played by the index grading in the cone (and cornea) was first appreciated by Exner⁹ as long ago as 1891. Recently Caveney and McIntyre¹⁰⁻¹¹ published beautiful studies of the ray paths in the superposition eyes of Scarab beetles.

The formalism of graded-index optics finds natural application in other, related disciplines. For example, when considering small amplitude disturbances, the propagation of acoustic waves beneath the surface of the ocean is completely analogous to the propagation of light in GRIN media. Here a generalized refractive index distribution follows from the varying sound velocity at different points in the ocean (which, in turn, results from the varying pressure, temperature and salinity)¹². Many effects studied in optics also occur naturally for acoustic waves within the ocean. For example, an underwater waveguide (referred to as an Underwater Sound Channel) results when the sound velocity has a minimum value at some depth beneath the surface. (This frequently occurs when the sound velocity increases above this depth due to increase in temperature, and increases below this depth due to the increase in hydrostatic pressure.) Sources of low frequency (for which absorption by sea water is small) can propagate power for distances measured in the thousands of kilometers¹².

Similar analogs exist in the fields of radiowave propagation¹³ and geophysics (elastic waves in the earth)¹⁴.

1.2 RECENT HISTORY OF GRIN OPTICS

There has been only minor interest in GRIN optics, until comparatively recently. Although the equations describing ray propagation in inhomogeneous media (Fermat's principle) have been known for many years, interest has been limited by the lack of successful fabrication techniques. This situation has changed only within the last two decades.

Historically, the most notable contributions were by Maxwell, Wood and Luneburg. Maxwell¹⁵, in 1854, described a graded-index lens of spherical symmetry (now referred to as Maxwell's Fisheye) with the interesting property that every ray leaving a point source inside (or on the surface) of the lens, is perfectly focused at a conjugate point. In 1905, Wood¹⁶ constructed a primitive cylindrical GRIN lens out of gelatin, and demonstrated how thin transverse sections, with flat endfaces, could form images. Luneburg¹⁷, in his classic book of 1944, further developed the theory of ray propagation in inhomogeneous media, and described a spherically symmetric GRIN lens (the Luneburg lens) which brings to a sharp focus any beam of parallel rays. (Generalizations of Luneburg's lens were introduced later - see Marchand¹⁸.)

Despite subsequent contributions by Adler¹⁹, and Fletcher, Murphy and Young²⁰, interest remained at a low level until the development of the laser, and the awareness of the enormous potential offered by lightwave communications. Most influential during this period was the work of Berreman²¹, Marcuse and Miller²² in 1964, and later other coworkers at Bell Labs and elsewhere who experimented with image and signal transmission through gas lenses. In its simplest form, a gas lens is formed by forcing a cool gas through a heated metal tube. The flow of heat creates a radial variation of gas density inside the tube, and consequently, a radial

refractive index distribution. A review of the experimental and theoretical work on gas lenses is given by Marcuse²³. Although only of historical interest now, the research into gas lenses demonstrated the feasibility of optical communications, and motivated development in other areas.

Also significant around this time was the theoretical analysis of propagation in transversely graded media, most notably by Tien et al.²⁴, Kogelnik²⁵, Gordon²⁶, Miller²⁷, Marcatili²⁸, Kawagami and Nishizawa²⁹, and Streiffer and Kurtz³⁰⁻³². The emphasis here was on the waveguiding properties of cylindrical GRIN media (for optical communications, and modelling the propagation of beams in laser media) as opposed to classical image formation.

An explosion of interest in GRIN optics followed the announcement of a successful technique for introducing significant index gradients in silica glass, by Hamblen³³ (1969) and in the landmark papers by Pearson et al.³⁴ (1969) and Uchida et al.³⁵ (1970). The technique, known as Ion Exchange, involves the replacement of alkali metal ions in the glass with metal ions of different size, by heating the glass in a molten salt bath. Rods with transverse grading can be obtained, with an approximately quadratic profile:

$$n(r) = n_0 \left\{ 1 - (1/2) r^2 / \rho^2 \right\} \quad (2)$$

(Here n_0 and ρ are constants, and (r, z, ϕ) are cylindrical coordinates with z defining the axis of the rod.) The appearance of these papers may be considered to mark the start of serious technological interest in graded-index optical devices. The paper by Uchida and coworkers, in

particular, describes early experiments in image transmission (through single GRIN rods, and a matrix of rods), multiplexing of optical pulses, coupling of power to photodiode detectors, and endoscopy.

A considerable amount of theoretical work on image formation with GRIN lenses followed the advances in fabrication. Foremost has been the work on ray tracing and aberration theory by Marchand, Moore, Sands and Kapron, which is summarized in the well-known book by Marchand¹⁸. A wave optics theory of imaging has been given by Iga and coworkers³⁶⁻³⁷.

Concurrently, there has been rapid development of graded-index fibers for fiber-optic communications. The field of fiber optics experienced an explosion of interest at about the same time as GRIN optics. Here the breakthrough was made by Kapron and coworkers³⁸ with the realization of low-loss silica glass fibers which made long-distance lightwave communications economically feasible. In its simplest design, an optical fiber is a dielectric cylinder of high refractive index (the core) surrounded by a lower index dielectric (the cladding). This is depicted schematically in Fig. 1(a).[†] A finite number of bound modes can propagate along the axis of the fiber, each with field concentrated in the core, but penetrating a little way into the cladding³⁹. When ray optics is applicable the guidance of light by the fiber may be viewed as arising from the total internal reflection of rays at the core-cladding interface (Fig. 1(b)). When the core refractive index is graded in the transverse direction (see Fig. 1(c)) the fiber continues to guide light, but the rays are now bent into curved trajectories and may turn around before reaching

[†] This figure is reproduced, with permission, from:
R.J. Black and A. Ankiewicz, *Am. J. Phys.* 53, 554 (1985).

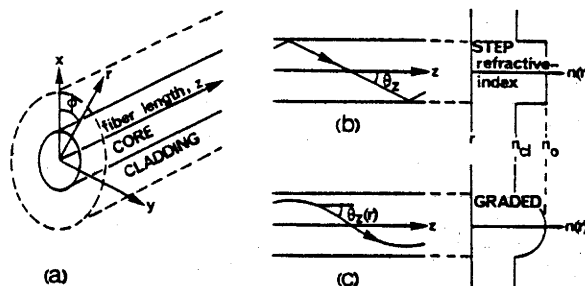


Fig. 1. (a) Optical fiber: the fiber core is embedded in a cladding of lower refractive index which, for theoretical purposes, one may effectively consider as extending to infinite r . Light ray trajectories in (b) step-index, and (c) graded-index fibers.

the interface (a point of much importance when imperfections at the core-cladding interface can cause scattering loss). An important distinction is made between fibers which support many bound modes, and those which can propagate only one or a few bound modes. (The specific number of guided modes depends on the value of the waveguide parameter, v ^{23,39}.) The former are referred to as multimoded, and can be used for the transmission of images (the number of bound modes is related to the image quality⁴⁰). The methods of ray optics may be utilized to study the behaviour of multimode waveguides. The latter are referred to as single- or few-moded, and are most suitable for long-distance communications (the information carrying capacity of a transmission line is limited by pulse spreading, and if a source excites several modes then each will generally propagate with a different group velocity, thus contributing to the pulse dispersion^{23,39}). Note that ray optics is not adequate to describe the behavior of single- and few-mode fibers.

1.3 APPLICATIONS OF GRIN - NOW AND FUTURE

Since the pioneering work of Hamblen, Pearson, Uchida and coworkers, further developments in the manufacture of index gradients have occurred. Ion exchange remains the most widely used technique, but several other methods hold great technological promise. Of particular interest is the work of Moore⁴¹, Ohtsuka⁴² and Iga et al.⁴³ on plastic focusing fibers, and the use of Chemical Vapor Deposition to obtain small diameter Graded-Index fibers for telecommunications⁴⁴⁻⁴⁵.

As gradient index materials become more widely available, commercial applications for GRIN optics can be expected to grow. The first commercial success has been in the imaging optics of photocopy machines⁴⁶, where an array of GRIN rods is used to form a superposition image "in a manner much like the insect eye"⁴⁷. The resulting system is more compact than can be attained with conventional lens optics, with no loss in image quality.

Single GRIN rods are used in medical endoscopes (the first experiments were described by Uchida and coworkers³⁵ in 1970). The more conventional fiberscopes use a bundle of optical fibers arranged into a precisely ordered matrix (each fiber associated with a single resolution element in the image⁴⁰). The GRIN endoscope eliminates the difficult alignment problem by utilizing a single multimoded graded-index rod. Current research is on GRIN systems for holographic endoscopy⁴⁸.

A more recent application is in optical memory and laser disk systems. Here, GRIN lenses may be used to focus laser pulses onto compact disks for reading and writing information⁴⁹.

An area with considerable potential is the design of high quality imaging systems. By appropriate choice of the parameters defining the

index distribution of a GRIN lens, the designer can control certain aberrations and reduce the number of distinct components required in an optical system. Currently, there are few successful examples⁵⁰, and competition is expected from the manufacturers of high quality aspherics⁵¹.

Within the communications field, there has been much research into the design of optical devices for multimode fiber systems. All-optical connectors, attenuators, branching junctions and directional couplers, multiplexers and demultiplexers, switches and isolaters have all been successfully constructed using short GRIN rods (see, for example, the reviews by Kobyashi et al.⁵² and Tomlinson⁵³). New ideas are required to provide the same functions for single-mode fiber systems. This is because in multimode systems, the optical power is distributed among many guided modes, and any mode conversion following the insertion of one of the above devices will lead to a redistribution of power among the modes, often with negligible macroscopic effect. Single-mode systems, by contrast, are highly sensitive to mode conversion, since any conversion is necessarily to the radiation field and therefore leads to loss. Current research indicates that GRIN lenses may find application as low-loss connectors for single-mode fibers of different cross-section⁵⁴, and for coupling power from a laser diode into a single-mode fiber⁵⁵⁻⁵⁷.

A new approach to the design of optical devices has recently been suggested by Iga and coworkers^{45,58}. Their approach, Stacked Planar Optics, takes advantage of planar technology by producing GRIN lenses on a planar substrate. Optical devices are constructed by stacking the lens array with similar arrays of other components, also produced on planar substrates.

1.4 TAPERED GRIN MEDIA

A medium which is graded in the direction transverse to the direction of propagation of a beam of light, may provide guidance (see section 1.2). We have referred to such a medium as a GRIN rod (or cylindrically symmetric graded-index medium), and a typical example is given by the index distribution of Eq. (2) with propagation along the z axis. The translational invariance leads to a considerable simplification in behavior - there are a finite number of forward and backward propagating bound modes, in addition to a radiation field^{23,39}. In this thesis we are concerned with light propagation in GRIN media with a longitudinal index gradient (i.e. along the direction of propagation) in addition to the transverse gradient. We shall call this a tapered GRIN medium when the longitudinal gradient monotonically increases or decreases down the axis. (In this way we exclude from consideration, for example, periodic variations in the longitudinal gradient.)

The earliest experiment on tapered GRIN rods was reported by Uchida et al.³⁵ in 1970. They fabricated rods with index distribution

$$n(r,z) = n_0 \left\{ 1 - (1/2) r^2 / \rho^2(z) \right\} \quad (3)$$

(compare with Eq. (2)). Here the longitudinal variation is through the characteristic radius of the transverse grading, $\rho(z)$, which slowly increased along the length of the rod.

The loss of translational invariance leads to fascinating propagation characteristics which are quite unlike modal propagation in conventional GRIN rods. For example, light injected into a down-tapering GRIN rod is confined to a progressively smaller region until it eventually

radiates away, or is reflected out in the opposite direction. Some tapered GRIN media have guided wave solutions⁵⁹ (bound modes with curved wavefronts) whereas others will always radiate continuously along their length. In the regime in which many guided (or "approximately guided" local modes) can propagate, tapered GRIN media exhibit useful and interesting imaging properties (for example, image magnification and reduction).

For these reasons, the behavior of light in tapered GRIN media is interesting in its own right. But in addition to its intrinsic interest, there are also a number of promising applications. Mikaelian⁶⁰ has pointed out that tapered GRIN media are useful models for studying the self-focusing of pulses in laser media. Furthermore, tapered GRIN media offer the promise of new optical devices.

1.4.1 WHY TAPER GRIN RODS?

1. Tapering offers optical designers additional degrees of freedom for the control of aberrations in high quality imaging systems.
2. Tapered GRIN rods can transmit images with magnification or reduction (see Gomez-Reino et al.⁶¹, or chapter 6 of this thesis). This can be achieved in a fiber bundle by altering the separation between the fibers at either end, or slowly tapering each individual fiber in the matrix⁶². The alignment problems inherent in this approach are completely eliminated by using a single tapered GRIN rod.
3. Graded-Index tapers may find application as waveguide connectors for waveguides of different size cross-section. The matching of multimode

waveguides with tapered GRIN rods has been discussed by Sawa⁶³. However, low-loss connection of single-mode fibers of different cross-section remains a significant problem in optical communications. A recent proposal by Marcatili⁵⁹ involves the use of strictly adiabatic tapers - GRIN tapers with guided wave solutions - and is discussed in chapter 7.

4. Light Concentration. There is considerable interest in the design of highly efficient light concentrators for application with light detectors and solar collectors⁶⁴. In effect, the GRIN taper may be used to increase the light receiving area of a detector (e.g. a photodiode) placed at its narrow end. Early experiments were reported by Uchida et al.³⁵ and Nishida et al.⁶⁵, and early theoretical work by Marcuse⁶⁶. Graded-Index tapers are particularly well suited to this task because the detector may be simply glued to the taper endface, eliminating the alignment problems associated with conventional lenses and reflecting concentrators. The strictly adiabatic tapers proposed by Marcatili⁵⁹ may be of particular relevance here (see chapter 4 for further discussion).
5. An Application from Nature: GRIN tapers in vision. Cone shaped optical elements, with index grading, can be identified in the eyes of some invertebrates⁴. The crystalline cones found in the compound eye of the Butterfly, for example, have been the subject of recent investigations^{7,67}. The cone is a prominent feature of each ommatidium, where it feeds light from the cornea down to the photoreceptor (the optics here is apposition). Experiments carried out

by Nilsson and coworkers suggest that this cone is a strongly tapered graded-index rod, and a model which explains the unusual measured characteristics of this structure is presented in chapter 8.

It has also been suggested that tapered GRIN rods may be useful for modelling the cone receptors in the retina of the human eye⁶⁸, but this is highly speculative.

1.4.2 RECENT PROGRESS WITH THEORY

One of the earliest theoretical studies into the effect of tapering a GRIN medium appears to have been carried out by Suematsu and Kitano⁶⁹. (Longitudinal non-uniformities first received attention regarding gravitational distortions in gas lenses⁷⁰. However, in non-optics fields, work goes back even further⁷¹.) The subsequent fabrication of tapered self-focusing rods by Uchida and coworkers³⁵ together with the demonstration of useful applications⁶⁵, stimulated further interest with early work by Yamamoto and Makimoto⁷²⁻⁷³ on ray matrices, and by Sodha, Ghatak and Malik⁷⁴⁻⁷⁷.

The main thrust in theoretical effort has concerned the theory of mode coupling in tapered optical waveguides^{23,39}. There are several ways of formulating coupled-mode theory, but the most useful for taper applications involves the concept of the local mode: at each location down the (z) axis we associate an imaginary, straight waveguide, which matches the cross-section of the taper at that point. The local modes of the taper are the modes of the imaginary waveguide, at each position along the axis, so that the z-dependence appears via the changing parameters which define these modes. The local modes are clearly not exact solutions of the wave equation - in fact, the exact solution is a linear combination of all the

local modes (plus radiation field) with z -dependent modal amplitudes. These modal amplitudes satisfy a set of coupled differential equations (the coupled-mode equations) which describe the interchange of power between the local modes. Approximate or numerical solutions may be obtained in certain situations (e.g. weak power transfer)^{23,39}.

The adiabatic approximation³⁹ is used extensively in optical fiber theory. This may be viewed as a special case of coupled-mode theory which applies when the degree of tapering is so small that the local modes exchange a negligible amount of power, and so coupling between them can be neglected. In ray theory, the analogous situation leads to a quantity which is approximately conserved along a ray path - the adiabatic invariant^{39,78}. Coupled mode and adiabatic theories of propagation have been around for many years, motivated in the underwater acoustics field by the problem of wave propagation in an ocean with a sloping bottom¹².

The propagation of Gaussian beams in tapered GRIN media has been studied by Sawa⁶³ and Casperson and coworkers⁷⁹⁻⁸¹.

Also to be mentioned here are contributions by Sharma and Goyal⁶⁸ on constant V (waveguide parameter^{23,39}) tapers, several recent geometrical optics studies⁸²⁻⁸⁴, and the work on paraxial imaging theory by Gomez-Reino and coworkers⁶¹ (see also chapter 6).

Perhaps the most significant recent development concerns the theory of lossless GRIN tapers (referred to as strictly adiabatic tapers). A remarkable paper by Marcatili⁵⁹ has shown the existence of a special class of graded-index taper which exhibits guided-wave (i.e. bound mode) propagation. These tapers are found by seeking separable solutions to the wave equation in various orthogonal coordinate systems (of course, not all index distributions allowing separation of the wave equation are

guiding - criteria are discussed by Marcatili⁵⁹). Conventional, straight GRIN waveguides are obtained when cylindrical polar (or rectangular) coordinates are used, but other coordinate systems lead to waveguides with tapering profiles and curved axes. The separated fields represent propagation down the taper with unchanging field shape, and with no coupling to other solutions (hence the term, strictly adiabatic taper). The concept of the mode is easily generalized to include these new guided fields.

1.5 BACKGROUND THEORY: LIGHT PROPAGATION IN GRADED-INDEX MEDIA

Light is an electromagnetic wave, and consequently its behavior is properly described by Maxwell's equations. These equations describe the propagation of the two vector fields, the electric field $\underline{E}(\underline{R}, t)$ and the magnetic field $\underline{H}(\underline{R}, t)$, which together comprise an electromagnetic wave. A medium with graded refractive index is a dielectric with a permittivity that varies with position, $\epsilon = \epsilon(\underline{R})$. In practice, GRIN media are often non-magnetic (i.e. have magnetic permeability, μ , the same as for free-space, $\mu = \mu_0$) and have no free electric charges. We also assume, in this thesis, isotropic media. Thus ϵ is a scalar function, related to the refractive index through the expression^{23,39} $n^2(\underline{R}) = c^2 \mu_0 \epsilon(\underline{R}) = \epsilon(\underline{R}) / \epsilon_0$, where c is the vacuum speed of light, and ϵ_0 the vacuum permittivity.

Furthermore, for short sections of GRIN media it is often acceptable to neglect loss (usually valid for GRIN lenses, but inappropriate for modelling long-distance communications fibers for which even a minute amount of loss will be significant). Allowance for loss will be discussed later, but for the moment we neglect conduction currents and write Maxwell's equations:

$$\nabla \times \underline{H}(\underline{R}, t) = \epsilon \partial \underline{E} / \partial t \quad (4a)$$

$$\nabla \times \underline{E}(\underline{R}, t) = - \mu_0 \partial \underline{H} / \partial t \quad (4b)$$

$$\nabla \cdot (\epsilon \underline{E}) = 0 \quad (4c)$$

$$\nabla \cdot \underline{H} = 0 \quad (4d)$$

These equations may be simplified by decoupling the electric and magnetic vectors (see Marcuse²³) giving

$$\nabla^2 \underline{E} - \mu_0 \epsilon \partial^2 \underline{E} / \partial t^2 = - \nabla [\underline{E} \cdot (\nabla \epsilon) / \epsilon] \quad (5a)$$

$$\nabla^2 \underline{H} - \mu_0 \epsilon \partial^2 \underline{H} / \partial t^2 = [\nabla \times \underline{H}] \times (\nabla \epsilon) / \epsilon \quad (5b)$$

The above equations completely describe light propagation in isotropic, non-magnetic, GRIN media, with no free charges or conduction currents. It is understood in Eqs. (5) that a cartesian coordinate system be used²³.

In many situations of interest, the permittivity, ϵ , may be assumed to vary slowly with position. If the relative change in ϵ over a distance of a wavelength is much less than unity, then the terms on the RHS of Eqs. (5) can be dropped with negligible error (see, for example, Marcuse's²³ argument involving order of magnitude estimates of the terms in Eqs. (5)). Thus the problem is conveniently simplified to the behavior of waves described by the scalar wave equation:

$$\nabla^2 \Psi - (n^2/c^2) \partial^2 \Psi / \partial t^2 = 0 \quad (6)$$

where Ψ denotes any of the cartesian components of the electric or magnetic field vectors. For most practical situations this is an excellent approximation, but care must be taken when polarization effects are considered^{23,39}. In the context of optical waveguide theory, this approximation is known as the weakly guiding assumption^{39,45}.

Further simplification arises by considering only a single Fourier component of the time-varying field:

$$\Psi(\underline{R}, t) = \text{Re}\{\psi(\underline{R}) \exp(-i\omega t)\} \quad (7)$$

Here $\psi(\underline{R})$ is the complex-valued scalar field which is associated with the angular frequency $\omega = ck = 2\pi c/\lambda$ (λ denotes the free-space wavelength), and satisfies

$$\nabla^2 \psi + k^2 n^2 \psi = 0 \quad (8)$$

Equation (8) is the Helmholtz equation, and is used extensively in theoretical studies of light propagation in GRIN media.

These basic equations are modified when conduction loss is accounted for. This is formally carried out by including a conduction current term $\underline{J}_{\text{cond}} = \sigma \underline{E}$ (here σ is the conductivity) to Maxwell's equation (4a), and leads to the addition of a damping term $-\mu_0 \sigma \partial \Psi / \partial t$ to the scalar wave equation. Monochromatic fields, in the form of Eq. (7),

have spatial dependence $\psi(\underline{R})$ which still satisfies the Helmholtz equation, but with n^2 replaced by $n^2 + i\alpha_0 \sigma/k$ in Eq. (8). A medium with gain (the active region of a laser, for example) may also be described by a complex dielectric constant, but with an imaginary part of opposite sign ($\sigma < 0$)²³.

Ray optics may be derived from the Helmholtz equation by considering wavefronts and their orthogonal trajectories (rays) in the limit $k \rightarrow \infty$ (the geometrical optics limit)²³. The ray paths are obtained from the well-known equation:

$$\frac{d}{ds} \left(n \frac{d\underline{R}}{ds} \right) = \underline{\nabla} n \quad (9)$$

where $\underline{R}$ is a position vector tracing out the trajectory, and s denotes length measured along the path. Equation (9) may be obtained more directly from Fermat's principle (see, for example, Marchand¹⁸).

It is interesting to note that the path traced out by a ray in a GRIN medium, $n = n(\underline{R})$, is analogous to the motion of a particle of zero energy in an external potential $U = -n^2(\underline{R})/2$ (see Luneburg¹⁷). Formally, this arises as a consequence of both trajectories satisfying variational principles: the ray follows a path which gives a stationary value to the optical path length (Fermat's principle), whereas a particle follows a path that gives a stationary value to the mechanical action (Hamilton's principle). The ray invariants in optics (non-trivial quantities which are conserved along a ray path) correspond to the integrals of motion in mechanics.

Exact solutions of the Helmholtz equation (Eq. (8)) and the ray equation (Eq. (9)) can only be obtained in very special cases. Until now, the solutions given by Marcatili⁵⁹ have been the only exact, non-trivial studies of tapering GRIN media. Usual approximation methods involve a

slowly-varying envelope approximation in wave theory (see chapter 5) or the paraxial ray equation²³ in geometrical optics.

1.6 ABOUT THIS THESIS

This thesis is concerned with some aspects of the behavior of light propagating in tapered GRIN media. The emphasis is on guided-wave propagation, diffraction and self-focusing, and some applications of contemporary interest.

In chapters 2 and 3 we present exact ray analysis and wave theory for a whole class of graded-index media. A tapered quadratic-index medium is one member of this class, and guided propagation here is studied in detail. An interesting feature of the solutions is that they encompass the propagation, compression and reflection of light within the taper.

Marcatili's paper on strictly adiabatic tapers gives wave theory only. In chapter 4 we give an exact ray analysis of one of these tapers, compare properties of ray and field solutions, and obtain a clearer, more physical understanding of radiation-free guidance in these structures. In addition, the compression of light inside the taper is discussed from the viewpoint of concentrator theory⁶⁴.

Scalar diffraction theory is treated in chapters 5 and 6. A theory of diffraction in quadratic-index media, based on a mode expansion of the field, has appeared in the literature^{36,37,77}. In chapter 5 we further develop the theory, and present a detailed study of the 3D light distribution at the foci of a self-focused field, and compare with the classical problem of the focused field distribution in a uniform medium. In chapter 6 we extend the theory to the case of a slowly, but otherwise arbitrarily, tapered medium. The effect of the tapering on the 3D light

distribution, and on image transmission, is emphasized. The advantage of the approach presented here is its conceptual simplicity, and results are consistent with those obtained from a more cumbersome Green's function approach.

Two applications of the theory of earlier chapters are presented in chapters 7 and 8. Chapter 7 concerns a problem of interest in the fiber-optics field: the low-loss connection of waveguides of different cross-section. We analyse an approach using strictly adiabatic tapers as waveguide connectors, and quantify the loss for slab waveguides. In chapter 8 we present a model for the crystalline cone of the Butterfly compound eye. The attractive physical properties of the model are discussed, and we show how it explains the recent measurements by Nilsson and coworkers^{7,67}.

1.7 REFERENCES

1. A.B. Fraser and W.H. Mach, "Mirages", Scien. Am. pg. 102, Jan. 1976.
2. A.B. Fraser, "So You Think the Sun is Round," Earth and Min. Scien. 45, 49 (1976).
3. E. Hecht and A. Zajac, "Optics," (Addison-Wesley, Reading, Massachusetts, 1979).
4. M.F. Land, "Compound Eyes: Old and New Optical Mechanisms," Nature 287, 681 (1980).

5. C. Pask and K.F. Barrell, "Photoreceptor Optics I: Introduction to Formalism and Excitation in a Lens-Photoreceptor System," Biol. Cybernetics 36, 1 (1980).
6. C. Pask and K.F. Barrell, "Photoreceptor Optics II: Application to Angular Sensitivity and Other Properties of Lens-Photoreceptor System," Biol. Cybernetics 36, 9 (1980).
7. D.-E. Nilsson, M.F. Land, and J. Howard, "Afocal Apposition Optics in Butterfly Eyes," Nature 312, 561 (1984).
8. G.A. Horridge, "Optical Mechanisms of Clear-Zone Eyes," in The Compound Eye and Vision of Insects, ed. G.A. Horridge (Clarendon Press, Oxford, 1975).
9. S. Exner, "Die Physiologie der Facettirten Augen von Krebsen und Insecten," (Franz Deuticke, Leipzig-Wien, 1891).
10. S. Caveney and P. McIntyre, "Design of Graded-Index Lenses in the Superposition Eyes of Scarab Beetles," Phil. Trans. R. Soc. Lond. B294, 589 (1981).
11. P. McIntyre and S. Caveney, "Graded-Index Optics are Matched to Optical Geometry in the Superposition Eyes of Scarab Beetles," Phil. Trans. R. Soc. Lond. B311, 237 (1985).

12. L. Brekhovskikh and Yu. Lysanov, "Fundamentals of Ocean Acoustics," (Springer-Verlag, Berlin, 1982).

13. K.G. Budden, "The Propagation of Radio Waves: The Theory of Radio Waves of Low Power in the Ionosphere and Magnetosphere," (Cambridge University Press, Cambridge, 1985).

14. M. Ewing, W.S. Jardetzky and F. Press, "Elastic Waves in Layered Media," (McGraw-Hill, New York, 1957).

15. J.C. Maxwell, "The Scientific Papers of James Clerk Maxwell," W.D. Niven, Ed. (Dover, New York, 1965).

16. R.W. Wood, "Physical Optics," (Macmillan, New York, 1905).

17. R.K. Luneburg, "Mathematical Theory of Optics," (University of California Press, Berkeley, 1964).

18. E.W. Marchand, "Gradient Index Optics," (Academic, New York, 1978).

19. R.B. Adler, "Waves on Inhomogeneous Cylindrical Structures," Proc. IRE 40, 339 (1952).

20. A. Fletcher, T. Murphy and A. Young, "Solutions of Two Optical Problems," Proc. Roy. Soc. A223, 216 (1954).

21. D.W. Berreman, "A Lens or Light Guide using Convectively Distorted Thermal Gradients in Gases," B.S.T.J. 43, 1469 (1964).
22. D. Marcuse and S.E. Miller, "Analysis of a Tubular Gas Lens," B.S.T.J. 43, 1759 (1964).
23. D. Marcuse, "Light Transmission Optics," (Van Nostrand Reinhold, New York, 1982).
24. P.K. Tien, J.P. Gordon and J.R. Whinnery, "Focusing of a Light Beam of Gaussian Field Distribution in Continuous and Periodic Lens-Like Media," Proc. IEEE 53, 129 (1965).
25. H. Kogelnik, "On the Propagation of Gaussian Beams of Light Through Lenslike Media Including those with a Loss or Gain Variation," Appl. Opt. 4, 1562 (1965).
26. J.P. Gordon, "Optics of General Guiding Media," B.S.T.J. 45, 321 (1966).
27. S.E. Miller, "Light Propagation in Generalized Lens-Like Media," B.S.T.J. 44, 2017 (1965).
28. E.A.J. Marcatili, "Off-Axis Wave-Optics Transmission in a Lens-Like Medium with Aberration," B.S.T.J. 46, 149 (1967).

29. S. Kawakami and J. Nishizawa, "An Optical Waveguide with the Optimum Distribution of the Refractive Index with Reference to Waveform Distortion," IEEE Trans. Microwave Theory and Techniques MTT-16, 814 (1968).
30. W. Streiffer and C.N. Kurtz, "Scalar Analysis of Radially Inhomogeneous Guiding Media," J. Opt. Soc. Am. 57, 779 (1967).
31. C.N. Kurtz and W. Streiffer, "Guided Waves in Inhomogeneous Focusing Media, Pt. I: Formulation, Solution for Quadratic Inhomogeneity," IEEE Trans. Microwave Theory and Techniques MTT-17, 11 (1967).
32. C.N. Kurtz and W. Streiffer, "Guided Waves in Inhomogeneous Focusing Media, Pt II: Asymptotic Solution for General Weak Inhomogeneity," IEEE Trans. Microwave Theory and Techniques MTT-17, 250 (1969).
33. D.P. Hamblen, "Gradient Refractive Index Optical Lenses," U.S. Patent 3,486,808 (1969).
34. A.D. Pearson, W.G. French and E.G. Rawson, "Preparation of a Light Focusing Glass Rod by Ion-Exchange Techniques," Appl. Phy. Lett. 15, 76 (1969).
35. T. Uchida, M. Furukawa, I. Kitano, K. Koizumi and H. Matsumura, "Optical Characteristics of Light-Focusing Fiber Guide and its Applications," IEEE J. Quantum Electronics QE-6, 606 (1970).

36. K. Iga, S. Hata, Y. Kato and H. Fukuyo, "Image Transmission by an Optical System with a Lens-Like Medium," Jpn. J. Appl. Phys. 13, 79 (1974).
37. K. Iga, "Theory for Gradient-Index Imaging," Appl. Opt. 19, 1039 (1980).
38. F.P. Kapron, D.B. Keck and R.D. Maurer, "Radiation Losses in Glass Optical Waveguides," Appl. Phys. Lett. 17, 423 (1970).
39. A.W. Snyder and J.D. Love, "Optical Waveguide Theory," (Chapman and Hall, London, 1983).
40. A.A. Friesem, U. Levy and Y. Silberberg, "Parallel Transmission of Images through Single Optical Fibers," Proc. IEEE 71, 208 (1983).
41. R.S. Moore, "Plastic Optical Element Having Refractive Index Gradient," U.S. Patent 3,718,383 (1973).
42. Y. Ohtsuka, "Light-Focusing Plastic Rod Prepared from Diallyl Isophthalate-Methyl Methacrylate Copolymerization," Appl. Phys. Lett. 23, 247 (1973).
43. K. Iga, N. Yamamoto and Y. Matsuura, in "CLEOS, 25-27 May 1976, Digest of Technical Papers," paper WD9. (Optical Society of America, Washington, D.C., 1976).

44. D.T. Moore, "Gradient-Index Optics: a Review," Appl. Opt. 19, 1035 (1980).
45. K. Iga, Y. Kobubun and M. Oikawa, "Fundamentals of Microoptics," (Academic Press, New York, 1984).
46. M. Kawazu and Y. Ogura, "Application of Gradient-Index Fiber Arrays to Copying Machines," Appl. Opt. 19, 1105 (1980).
47. L. Pugliese and D.T. Moore, "Gradient Index Optics," Photonics Spectra, March 1987, pg. 71.
48. G. von Bally, E. Brune and W. Mette, "Holographic Endoscopy with Gradient-Index Optical Imaging Systems and Optical Fibers," Appl. Opt. 25, 3425 (1986).
49. H. Nishi, H. Ichikawa, M. Toyama and I. Kitano, "Gradient-Index Objective Lens for the Compact Disk System," Appl. Opt. 25, 3340 (1986).
50. J.B. Caldwell, L.R. Gardner, S.N. Houde-Walter, M.T. Houk, D.S. Kindred, D.T. Moore, D.P. Naughton and C.B. Wooley, "Gradient Index Binocular Objective Design," Appl. Opt. 25, 3345 (1986).
51. R.S. Hilbert, "Gradient Index is Growing," Photonics Spectra Dec. 1982 pg. 59.

52. K. Kobayashi, R. Ishikawa, K. Minemura and S. Sugimoto, "Micro-Optic Devices for Fiber-Optic Communications," *Fiber Integrated Opt.* 2, 1 (1979).
53. W.J. Tomlinson, "Applications of GRIN-Rod Lenses in Optical Fiber Communications Systems 19, 1127 (1980).
54. E.-G. Neumann, "Beam Transformers for Obtaining Low-Loss Splices Between Dissimilar Single-Mode Fibers," *J. Opt. Soc. Am.* A4, 1021 (1987).
55. K. Kawano, "Coupling Characteristics of Lens Systems for Laser Diode Modules using Single-Mode Fiber," *Appl. Opt.* 25, 2600 (1986).
56. I. Kitano, H. Ueno and M. Toyama, "Gradient-Index Lens for Low-Loss Coupling of a Laser Diode to Single-Mode Fiber," *Appl. Opt.* 25, 3336 (1986).
57. H. Honmou, R. Ishikawa, H. Ueno and M. Kobayashi, "1.0 dB Low-Loss Coupling of Laser Diode to Single-Mode Fibre Using a Planoconvex Graded-Index Rod Lens," *Electronics Lett.* 22, 1122 (1986).
58. K. Iga and S. Misawa, "Distributed-Index Planar Microlens and Stacked Planar Optics: a Review of Progress," *Appl. Opt.* 25, 3388 (1986).
59. E.A.J. Marcatili, "Dielectric Tapers with Curved Axes and No Loss," *IEEE J. Quantum Electron.* QE-21, 307 (1985).

60. A.L. Mikaelian, "Self-Focusing Media with Variable Index of Refraction," Prog. Opt. XVII, 279 (1980).
61. C. Gomez-Reino and E. Larrea, "Imaging and Transforming Transmission Through a Medium with Nonrotation-Symmetric Gradient-Index," Appl. Opt. 22, 387 (1983).
62. N. Takato, T. Kurokawa and Y. Katayama, "Tapered Fiber Optic Sheet Formation and its Image-Guiding Properties," Appl. Opt. 20, 1397 (1981).
63. S. Sawa, "A Theoretical Study of Light Beams Guided Along Tapered Lenslike Media, and Their Applications," IEEE Trans. Microwave Theory Tech. MTT-24, 77 (1976).
64. W.T. Welford and R. Winston, "The Optics of Nonimaging Concentrators," (Academic Press, New York, 1978).
65. K. Nishida, Y. Nannichi, T. Uchida and I. Kitano, "An Avalanche Photodiode with a Tapered Light-Focusing Fiber Guide," Proc. IEEE 58, 790 (1970).
66. D. Marcuse, "Compression of a Bundle of Light Rays," Appl. Opt. 10, 494 (1971).
67. D.-E. Nilsson, M.F. Land and J. Howard, "Optical Characteristics of the Butterfly Eye," J. Comp. Physiol. (in press).

68. A. Sharma and I.C. Goyal, "Tapered Parabolic-Index Fiber Waveguide," Appl. Opt. 18, 1482 (1979).
69. Y. Suematsu and T. Kitano, "Analysis of a Tapered Lens-Like Beam Guide," IECE Nat. Conv. Rec. (Japan), No. 535 (1967).
70. W.H. Steier, "Measurements on a Thermal Gradient Gas Lens," IEEE Trans. Microwave Theory Tech. MTT-13, 740 (1965).
71. D.E. Weston, "Guided Propagation in a Slowly Varying Medium," Proc. Phys. Soc. London 73, 365 (1959).
72. S. Yamamoto and T. Makimoto, "Equivalence Relations in a Class of Distributed Optical Systems - Lenslike Media," Appl. Opt. 10, 1160 (1971).
73. S. Yamamoto and T. Makimoto, "On the Ray Transfer Matrix of a Tapered Lenslike Medium," Proc. IEEE 59, 1254 (1971).
74. M.S. Sodha, A.K. Ghatak and D.P.S. Malik, "Electromagnetic Wave Propagation in Radially and Axially Nonuniform Media: Geometrical Optics Approximation," J. Opt. Soc. Am. 61, 1492 (1971).
75. M.S. Sodha, A.K. Ghatak and D.P.S. Malik, "Electromagnetic Wave Propagation in a Radially and Axially Non-Uniform Medium: Optics of SELFOC Fibers and Rods; Wave Optics Considerations," J. Phys. D. 4, 1887 (1971).

76. A.K. Ghatak, D.P.S. Malik and M.S. Sodha, "Path of Rays in Conical Selfoc Fibers," J. Opt. Soc. Am. 62, 594 (1972).
77. M.S. Sodha and A.K. Ghatak, "Inhomogeneous Optical Waveguides," (Plenum Press, New York, 1977).
78. A. Ankiewicz, C. Pask and A.W. Snyder, "Slowly Varying Optical Fibers," J. Opt. Soc. Am. 72, 198 (1982).
79. L.W. Casperson, "Gaussian Light Beams in Inhomogeneous Media," Appl. Opt. 12, 2434 (1973).
80. L.W. Casperson and J.L. Kirkwood, "Beam Propagation in Tapered Quadratic Index Waveguides: Numerical Solutions," J. Lightwave Tech. LT-3, 256 (1985).
81. L.W. Casperson, "Beam Propagation in Tapered Quadratic-Index Waveguides: Analytical Solutions," J. Lightwave Tech. LT-3, 264 (1985).
82. K. Thyagarajan, A. Rohra and A.K. Ghatak, "Aberrations in Conically Graded Index Media," Opt. Acta 26, 863 (1979).
83. S.J.S. Brown, "Geometrical Optics of Tapered Gradient-Index Rods," Appl. Opt. 19, 1056 (1980).
84. J.N. McMullin, "The ABCD Matrix in Arbitrarily Tapered Quadratic-Index Waveguides," Appl. Opt. 25, 2184 (1986).

CHAPTER 2

PROPAGATION IN A TAPERED GRADED-INDEX MEDIUM I:

EXACT RAY ANALYSIS

2.1 INTRODUCTION

Exact ray solutions can be valuable in clarifying the physical characteristics of light propagation, and may be of considerable assistance to optical designers. The rarity of exact solutions for tapering media has already been pointed out in the introduction to this thesis (section 1.5) and is the prime motivation for the work presented in this chapter.

Although tapered graded-index media are our main concern, we approach the problem by finding exact ray solutions for a whole class of graded-index media of which the taper is just one example. This special class still encompasses an enormous number of fundamentally different profiles, but the ray trajectories have certain features common to all members.

Following this introduction, we begin, in section 2.2, with a description of the general class of GRIN media to be analysed. After setting up the formalism for the problem in section 2.3, we go on to solve exactly for the three ray trajectory components, in the general case, in sections 2.4, 2.5 and 2.6. The characteristics of the ray trajectories which are common to all members of the class are elucidated in section 2.7.

In section 2.8 we apply the analysis of the previous sections to the detailed study of the ray paths in a tapered quadratic-index medium. Particular interest here is paid to the meridional ray solutions. The main

features of the paths are deduced and illustrated in section 2.9, and concluding remarks form section 2.10.

2.2 THE GRADED-INDEX MODEL

Referring to Fig. 1, we consider an x-y-z rectangular coordinate system, with the z-axis taken to be the axis of the optical system. We next define spherical polar coordinates with origin at $x = 0, y = 0, z = -D$ as shown in the figure (D is taken to be a positive constant).

The general class of refractive index distributions is defined by

$$n^2 = n_0^2 \{1 - (2\Delta)G(\theta)/R^2\} \quad , \quad (1)$$

where G is an arbitrary 2π -periodic function of θ , and Δ is a constant.

Of most interest in this chapter is the particular index distribution obtained by setting

$$G(\theta) = \sin^2\theta/\cos^4\theta \quad . \quad (2)$$

In the rectangular coordinates this can clearly be seen to correspond to a tapered quadratic-index medium:

$$n^2 = n_0^2 \{1 - (2\Delta) (x^2 + y^2)/(D + z)^4\} \quad . \quad (3a)$$

Figure 2 shows some constant-index contours for this distribution. We can also write this refractive index in a form which may be more familiar to some readers:

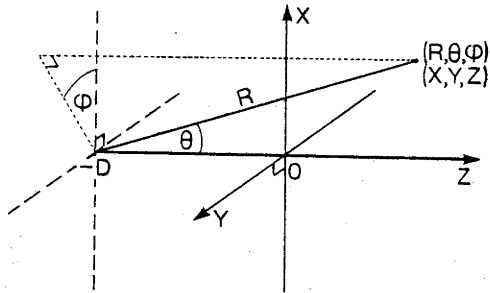


Fig. 1. Coordinate system. The origin of the R - θ - φ spherical coordinate system is located at point $x = 0, y = 0, z = -D$ of the rectangular system. We follow the usual convention: $0 \leq R < \infty, 0 \leq \theta \leq \pi$, and $0 \leq \varphi \leq 2\pi$.

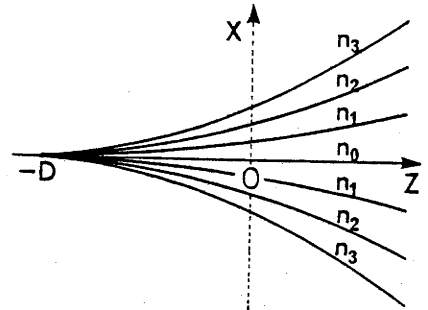


Fig. 2. Constant refractive-index contours for the tapered graded index medium defined by Eq. (3). Index values are ordered $n_3 < n_2 < n_1 < n_0$.

$$n^2 = n_0^2 \{1 - (2\delta)(x^2 + y^2)/\rho^2(z)\} \quad (3b)$$

where ρ , the characteristic radius for the grading in the transverse direction, is given by $\rho(z) = \rho_0(1 + z/D)^2$ and $2\delta = 2\Delta\rho_0^2/D^4$ is a dimensionless parameter. The medium can be described as an infinite quadratic-index medium with parabolic tapering.

By way of comparison, if we choose for the function G

$$G(\theta) = 1/\cos^2\theta, \quad (4a)$$

or

$$G(\theta) = 1, \quad (4b)$$

we obtain, respectively, media stratified in the z -direction or radially:

$$n^2 = n_0^2 \{1 - (2\Delta)/(D + z)^2\}, \quad (5a)$$

$$n^2 = n_0^2 \{1 - (2\Delta)/R^2\} \quad . \quad (5b)$$

Thus, although we concentrate particularly on tapered graded media as in Eq. (3), many varied profiles are contained within the general class defined by Eq. (1).

It should be stated at this point that profiles written in the form of Eq. (1) become unphysical for large values of $G(\theta)/R^2$. However, provided the ray trajectories lie entirely within the region

$$2\Delta G(\theta)/R^2 \ll 1 \quad (6)$$

this is of no concern, and in practice Eq. (6) will often be satisfied. If we concern ourselves with rays in the $z > 0$ half space, and D is sufficiently large, we may expect Eq. (6) to hold.

2.3 THE FORMALISM

2.3.1 THE RAY EQUATIONS

Let S be the length measured along a ray path, and define the parameter t by

$$dt = dS/n \quad . \quad (7)$$

Then the ray path is given by the equation¹:

$$\ddot{x} = d^2x/dt^2 = (1/2) \partial n^2 / \partial x \quad (8)$$

and similarly for y and z . (We use the convention of dotting a quantity to indicate d/dt .)

With this parameterization, the ray equations take the form of classical dynamics equations with potential $-n^2/2$ and zero energy (as noted in section 1.5) since elementary geometry gives

$$\dot{x}^2 + \dot{y}^2 + \dot{z}^2 - n^2 = 0 \quad . \quad (9)$$

Converting to spherical polars, we find

$$\ddot{R} - R\dot{\theta}^2 - R\dot{\phi}^2 \sin^2\theta = (1/2) \partial n^2 / \partial R \quad (10)$$

$$R\ddot{\phi} \sin\theta + 2R\dot{\theta}\dot{\phi} \cos\theta + 2R\dot{\phi} \sin\theta = (1/2 R \sin\theta) \partial n^2 / \partial \phi \quad (11)$$

and Eq. (9) becomes

$$\dot{R}^2 + R^2\dot{\theta}^2 + R^2 \sin^2\theta \dot{\phi}^2 - n^2 = 0 \quad . \quad (12)$$

For the class of refractive index profiles defined by Eq. (1), we can use Eqs. (10), (11) and (12) to obtain

$$\ddot{R}R = n_o^2 - \dot{R}^2 \quad , \quad (13)$$

$$R^4\dot{\theta}^2 = n_o^2 R^2 - (RR)^2 - w^2 / \sin^2\theta - 2\Delta n_o^2 G(\theta) \quad , \quad (14)$$

$$w = R^2\dot{\phi} \sin^2\theta \quad . \quad (15)$$

Here w is an invariant for each ray path (sometimes called the "skew invariant"²) and appears whenever a profile is rotation symmetric.

2.3.2 INITIAL CONDITIONS

Equations (13) to (15) contain the "dynamics" of the ray trajectories. They can be solved to obtain R , θ and ϕ explicitly in terms of the t -parameter, once the initial conditions are specified.

In terms of spherical polar coordinates, the initial conditions required are the ray position

$$\begin{aligned} R(t = 0) &= R_1 \\ \theta(t = 0) &= \theta_1 \\ \phi(t = 0) &= \phi_1 \end{aligned} \tag{16a}$$

and any two of the direction cosines

$$C_x = \cos(\mu_x), \quad C_y = \cos(\mu_y), \quad C_z = \cos(\mu_z) \tag{16b}$$

relative to the x, y, z axes, respectively.

It then follows that

$$\dot{R}(t = 0) = \dot{R}_1 = n(R_1, \theta_1, \phi_1) \{ C_x \cos \phi_1 \sin \theta_1 + C_y \sin \phi_1 \sin \theta_1 + C_z \cos \theta_1 \} \tag{17a}$$

$$\dot{\theta}(t = 0) = \dot{\theta}_1 = (n(R_1, \theta_1, \phi_1) / R_1) \{ C_x \cos \phi_1 \cos \theta_1 + C_y \sin \phi_1 \cos \theta_1 - C_z \sin \theta_1 \} \tag{17b}$$

$$\dot{\phi}(t = 0) = \dot{\phi}_1 = (n(R_1, \theta_1, \phi_1)/R_1 \sin \theta_1) \{C_Y \cos \phi_1 - C_X \sin \phi_1\} \quad (17c)$$

2.4 RAY PATH: R VARIABLE

Re-writing Eq. (13) as

$$\frac{d}{dt} (R\dot{R}) = n_O^2 \quad (18)$$

we can integrate immediately to obtain the R-variable:

$$R(t) = \{n_O^2 t^2 + 2R_1 \dot{R}_1 t + R_1^2\}^{1/2} \quad (19)$$

where R_1 and $\dot{R}_1$ are the initial conditions. Examination of Eqs. (18) or (19) reveals that

$$R^2(n_O^2 - \dot{R}^2) = R_1^2(n_O^2 - \dot{R}_1^2) \equiv \kappa \quad (20)$$

Thus we have obtained a second invariant, κ , for each ray path. This invariant has a simple geometric interpretation in terms of β , the angle between the ray and the radial direction, as shown in Fig. 3. Equations (20) and (7) and elementary geometry give

$$\kappa = R^2(n_O^2 - n^2 \cos^2 \beta) \quad (21a)$$

If the subscript a denotes quantities on the optical axis (Fig. 3) and $G(0) = 0$ (i.e. refractive index on axis is n_O , as is the case when $G(\theta)$ is given by Eq. (2)) then

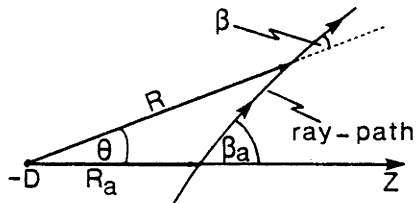


Fig. 3. Ray angle β at the point (R, θ) on a ray path. When the ray crosses the axis $\theta = 0$, $R = R_a$, and β_a indicates the steepness of the crossing.

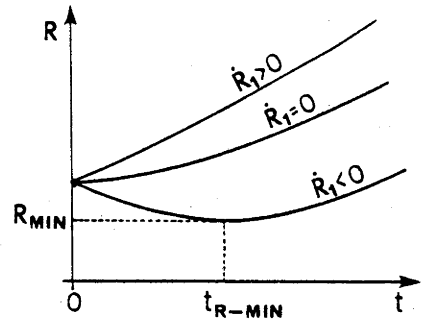


Fig. 4. Examples of R as a function of the t parameter. The curves are labeled by three possible initial condition cases: $\dot{R}_1 > 0$, $\dot{R}_1 < 0$, and $\dot{R}_1 = 0$.

$$\kappa = R_a^2 n_o^2 \sin^2 \beta_a, \quad (21b)$$

which gives a simple relationship between axis crossing position and ray steepness.

Curves showing R as a function of t are depicted in Fig. 4. Clearly, if $\dot{R}_1 > 0$ then R steadily increases as t increases from zero. However, if $\dot{R}_1 < 0$ then R first decreases as t increases from zero, reaches a minimum,

$$R_{\min} = \sqrt{\kappa} / n_o \quad (22)$$

at the value of t :

$$t_{R-\min} = -R_1 \dot{R}_1 / n_o^2 \quad (23)$$

and then steadily increases with t .

Thus, the solution for R is completely determined and reveals a very simple behaviour in terms of approach towards and recession from the origin $R = 0$.

2.5 RAY PATH: θ VARIABLE

We now turn to the θ -variable, using Eqs. (14) and (20) to obtain

$$R^4 \dot{\theta}^2 = \kappa - (w^2/\sin^2\theta) - 2\Delta n_O^2 G(\theta) \quad (24a)$$

$$\equiv \kappa \{1 - H(\theta)\} \quad (24b)$$

Observing that the R.H.S. of Eq. (24) can never be negative, we see that on the physical ray path

$$H(\theta) \equiv w^2/(\kappa \sin^2\theta) + \gamma^2 G(\theta) \leq 1 \quad (25)$$

where

$$\gamma = \sqrt{2\Delta n_O^2/\kappa} \quad (26)$$

Equation (25) requires that $\theta(t)$ be confined to lie between certain limits which depend upon the exact form of $G(\theta)$, and on the launching conditions of the ray:

$$\theta_{\min} \leq \theta(t) \leq \theta_{\max} \quad (27)$$

Here $\theta_{\min}$ and $\theta_{\max}$ are solutions of the equation

$$w^2/(\kappa \sin^2 \theta) + \gamma^2 G(\theta) = 1 \quad . \quad (28)$$

The θ -variable is obtained implicitly as a function of t by solving Eq. (24). This gives the result:

$$\int \frac{d\theta}{\sqrt{\{1 - w^2/(\kappa \sin^2 \theta) - \gamma^2 G(\theta)\}}} = \pm \tan^{-1} \{ (n_0^2 t + R_1 \dot{R}_1) / \sqrt{\kappa} \} \quad . \quad (29)$$

Unlike the R -variable case, this depends very definitely on the form of $G(\theta)$. Later we obtain $\theta(t)$ explicitly for the tapered quadratic index medium.

2.6 RAY PATH: ϕ VARIABLE

Finally, the ϕ -variable is obtained from Eq. (15), once $\theta(t)$ has been found:

$$\phi = w \int \frac{dt}{R^2(t) \sin^2 \theta(t)} \quad (30)$$

Rays launched with $w = 0$ are meridional (i.e. lie on a plane passing through the z -axis), whereas those rays with $w \neq 0$ are skew rays. If $w < 0$ then ϕ is a monotonically decreasing function of t everywhere, and conversely if $w > 0$. Thus the sign of w merely indicates whether the ray rotates clockwise or anticlockwise about the z -axis.

2.7 GENERAL FEATURES OF THE RAY PATHS

The above sections give the exact mathematical analysis of the ray trajectories, and we now turn to the qualitative features.

Firstly, once a ray is launched its R -variable either monotonically increases, or it decreases and reaches a minimum $R_{\min}$, given by Eq. (22), and then monotonically increases.

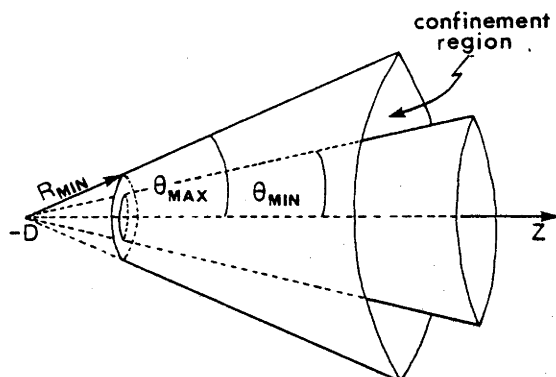


Fig. 5. Region of confinement of a ray.

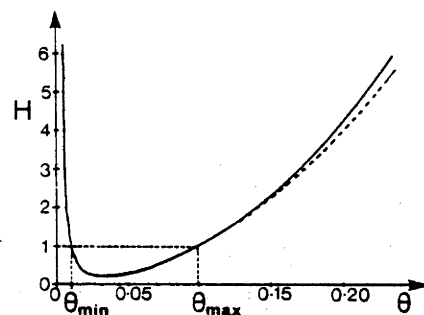


Fig. 6. Plot of $H(\theta)$, as defined in Eq. (24b) or (31), vs θ for sample values $w/\sqrt{\kappa} = 0.01$ and $\gamma = 10$. The condition $H(\theta) = 1$ defines the limiting angles $\theta_{\min}$ and $\theta_{\max}$ and physical ray paths only correspond to $H(\theta) \leq 1$. The dashed curve shows the small-angle approximation, $H(\theta) = w^2/\kappa\theta^2 + \gamma^2\theta^2$.

Secondly, the θ -variable is restricted to lie within the limits $\theta_{\min}$ and $\theta_{\max}$ (the exact values of which depend very definitely upon the specific form of $G(\theta)$).

Thus the ray is confined to lie within the region of space between two cones with half-angles $\theta_{\min}$ and $\theta_{\max}$, and bounded at the narrow end by the spherical surface $R = R_{\min}$. This region is illustrated in Fig. 5. (Note that since $R_{\min}$, $\theta_{\min}$ and $\theta_{\max}$ all depend on the initial conditions of the ray, the region of confinement will be different for different rays.)

We can picture the ray spiralling about the z -axis, within its region of confinement, as it propagates towards and then away from the origin, $R = 0$, with its closest approach at $R_{\min}$. However, the complete picture must also include details of the behavior of the θ -variable at the same time that this occurs. If, for example, the initial condition is $\dot{\theta}_1 > 0$, then θ must increase. But θ cannot increase beyond $\theta_{\max}$, so one of two things must happen; θ might tend asymptotically towards a

limiting value within the region of confinement, or θ will eventually reach $\theta_{\max}$. At the instant when the latter occurs, $\dot{\theta} = 0$ and beyond this $\dot{\theta} < 0$ (i.e. we switch from the positive to the negative $\dot{\theta}$ solution in Eq. (24)). The former case corresponds to the ray simply appearing to be "scattered" off the $R = R_{\min}$ surface, whereas the latter results in a complete oscillatory motion with the ray appearing to be "reflected" off the walls of the region of confinement (caustic surfaces, see chapter 3).

It is obvious that letting t take on negative values, i.e. considering $-\infty < t < \infty$, enables us to trace out the whole ray trajectory. We next turn to the specific example of a tapered quadratic-index medium.

2.8 RAY SOLUTION FOR THE GRIN TAPER

We study in detail the ray trajectories in the tapered graded-index medium defined by Eq. (3).

Figure 6 shows the function

$$H(\theta) = w^2 / (\kappa \sin^2 \theta) + \gamma^2 \sin^2 \theta / \cos^4 \theta \quad (31)$$

for typical parameter values. According to Eqs. (24) and (25) this function determines the behavior of $\theta(t)$ on the ray path. The requirement $H(\theta) \leq 1$ leads to the θ -variable limits, $\theta_{\min}$ and $\theta_{\max}$, as depicted in the figure.

We would like to express the θ and ϕ variables explicitly in terms of the t -parameter, but we are impeded by the inability to evaluate the θ -integral of Eq. (29) in closed form. However, if we focus attention on the small- θ region, which will correspond to many situations of interest, then we can make the approximations

$$\begin{aligned}\sin\theta &\approx \theta, \\ \cos\theta &\approx 1.\end{aligned}\tag{32}$$

Using Eq. (29) we obtain θ explicitly as a function of t :

$$\theta^2(t) = (\theta_{\max}^2 + \theta_{\min}^2)/2 + \{(\theta_{\max}^2 - \theta_{\min}^2)/2\} \sin\Omega\tag{33a}$$

where

$$\begin{aligned}\Omega = \pm 2\gamma \{ &\tan^{-1}[(n_0^2 t + R_1 \dot{R}_1)/\sqrt{\kappa}] - \tan^{-1}[R_1 \dot{R}_1/\sqrt{\kappa}]\} \\ &+ \sin^{-1}[(2\gamma^2 \theta_1^2 - 1) / (2\gamma^2 \theta_{\max}^2 - 1)]\end{aligned}\tag{33b}$$

The confining angles (determined from $H(\theta) = 1$) are given by

$$\theta_{\max}^2 = \{1 + \sqrt{(1 - 4\gamma^2 w^2/\kappa)}\}/2\gamma^2\tag{34}$$

$$\theta_{\min}^2 = \{1 - \sqrt{(1 - 4\gamma^2 w^2/\kappa)}\}/2\gamma^2\tag{35}$$

(Note, in Eq. (33), that $+$ or $-$ is used according to whether $\dot{\theta}_1 >$ or < 0 . Also note that throughout this thesis, $\sin^{-1}$ and $\tan^{-1}$ will be used to denote the principal branches of the inverse sine and tangent functions.)

The small- θ approximation will clearly be valid provided $\theta_{\max} \ll 1$. In Fig. 6 we have plotted the small-angle approximation to

$H(\theta)$ in order to demonstrate the validity of this approximation.

To determine the ϕ -variable, it proves more convenient to use the general equation

$$\frac{d\phi}{d\theta} = \pm \frac{w}{\sin^2\theta \sqrt{\kappa[1 - \gamma^2 G(\theta) - w^2/(\kappa \sin^2\theta)]}} \quad (36)$$

obtained from Eqs. (15) and (24). Equation (36) may be integrated to obtain ϕ as a function of θ . Within the small- θ approximation, the result is:

$$\phi(\theta) = \pm (1/2) \sin^{-1}\{(\theta^2 - 2w^2/\kappa)/\theta^2\sqrt{1 - 4\gamma^2 w^2/\kappa}\} + A \quad (37)$$

where A is a constant. Substituting $\theta(t)$ into Eq. (37) gives ϕ explicitly as a function of the t -parameter.

Thus, the explicit forms of all components of the ray trajectories are known: exactly for the R variable, Eq. (19), and in the small- θ approximation for $\theta(t)$ and $\phi(t)$.

2.8.1 MERIDIONAL RAYS

The meridional rays are an important sub-set of the total ray population. They are launched by a symmetric input to the taper, have skew invariant $w = 0$, and are confined to a plane. Since there is no ϕ variation for this case, we revert to plane-polar coordinates (R, θ) and allow θ to take on negative values.

(We also point out that a block of material uniform in the y -direction, tapered in the z -direction, and graded in the x -direction, has

rays which are described by the meridional paths plus the result $y(t) = \dot{y}(0)t + y(0)$.)

Integration of Eq. (29) with $w = 0$ gives the particularly simple result (in the small-angle approximation):

$$\begin{aligned} \theta(t) = \theta_{\max} \sin\left\{\pm \left(1/\theta_{\max}\right) \left[\tan^{-1}\left\{\left(n_0^2 t + R_1 \dot{R}_1\right)/\sqrt{\kappa}\right\} - \tan^{-1}\left\{R_1 \dot{R}_1/\sqrt{\kappa}\right\}\right] \right. \\ \left. + \sin^{-1}\left(\theta_1/\theta_{\max}\right)\right\} \end{aligned} \quad (38)$$

where $\pm$ is according to whether the initial value $\dot{\theta}_1 >$ or < 0 . The amplitude of the ray oscillation is given by

$$\theta_{\max} = 1/\gamma = \left\{\kappa/2\Delta n_0^2\right\}^{1/2} . \quad (39)$$

We explore the physical characteristics of these solutions in the following section.

2.9 BEHAVIOR OF RAYS IN THE TAPER

The nature of the meridional ray paths is more readily discussed if we use Eqs. (19) and (38) to eliminate t , and write θ as a function of R :

$$\theta = \theta_{\max} \sin\left\{\left(1/\theta_{\max}\right) \tan^{-1}\left[\pm \sqrt{(R^2 - R_{\min}^2)}/R_{\min}\right] + B\right\} . \quad (40)$$

The ray path constant κ is defined in Eq. (20), and $R_{\min} = \sqrt{\kappa}/n_0$. The constant B is fixed by any point on the ray path, e.g. our initial conditions at R_1, θ_1 . In particular, if $\theta = \theta_0$ when $R = R_{\min}$, then

$$B = \sin^{-1}(\theta_0/\theta_{\max}) \quad . \quad (41)$$

Note that R varies from $R_{\min}$ to ∞ , and using the two signs in Eq. (40) traces out the whole path, e.g. use $+$ sign for ∞ down to $R_{\min}$, then $-$ sign for $R_{\min}$ out to ∞ again. In practice energy may flow along this path in either direction as is clear from the t -parameterization.

The simple, explicit formula given in Eq. (40) allows us to extract all the features of ray behavior in this tapered medium. Obviously the path is confined to the region $-\theta_{\max} < \theta < \theta_{\max}$ and $R > R_{\min}$, and an oscillatory function is involved. However, the argument of the sine function is not just a linear function of R , but involves the inverse tangent function, as sketched in Fig. 7. From this we see that the argument has a total variation of $\pi/\theta_{\max}$ and so the whole path includes about $(\pi/\theta_{\max})/2\pi = 1/2\theta_{\max}$ periods, and the number of axis crossings N_a is

$$N_a \approx 1/2\theta_{\max} \quad . \quad (42)$$

We also know that the ray path constant, κ , indicates that the smaller the value of R , the steeper the angle when crossing the axis - see Eq. (21b). Setting $\theta = 0$ in Eq. (40) gives a formula for R_a , and from this, or just from consideration of the argument, Fig. 7, we see that axis crossing are closer together for smaller R .

When R becomes very large the $\tan^{-1}$ function approaches a limit (see Fig. 7) and so θ also tends to asymptotic values $\theta_{\pm\infty}$ given by

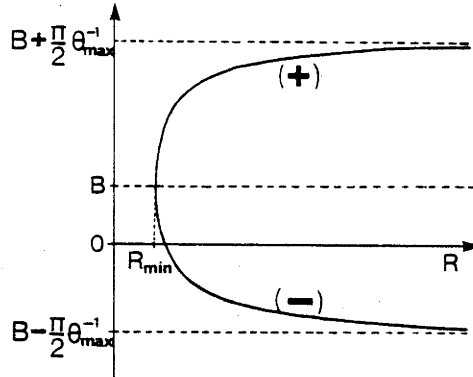


Fig. 7. Argument of the sine function in Eq. (40) giving the variation of θ with R . Sections of the curve labeled (+) or (-) correspond to taking those signs in Eq. (40).

$$\theta_{\pm\infty} = \theta_{\max} \sin\left\{\pm \pi / (2\theta_{\max}) + B\right\} \quad (43)$$

Using Eq. (41) we see that all the important angles are linked by

$$(\theta_{+\infty} + \theta_{-\infty}) = 2\theta_0 \cos(\pi / 2\theta_{\max}) \quad (44)$$

The main ray path features are shown semi-schematically in Fig. 8(a). Interesting special cases occur when:

$$\begin{aligned} \theta_{\max} = 1/2m, \quad \theta_{+\infty} = \theta_{-\infty} & \quad (\text{perfect reflector}) \\ \theta_{\max} = 1/(2m+1), \quad \theta_{+\infty} = -\theta_{-\infty} & \quad (\text{symmetric path when } \theta_0 = 0) \end{aligned} \quad (45)$$

and these are illustrated in Fig. 8(b) and (c).

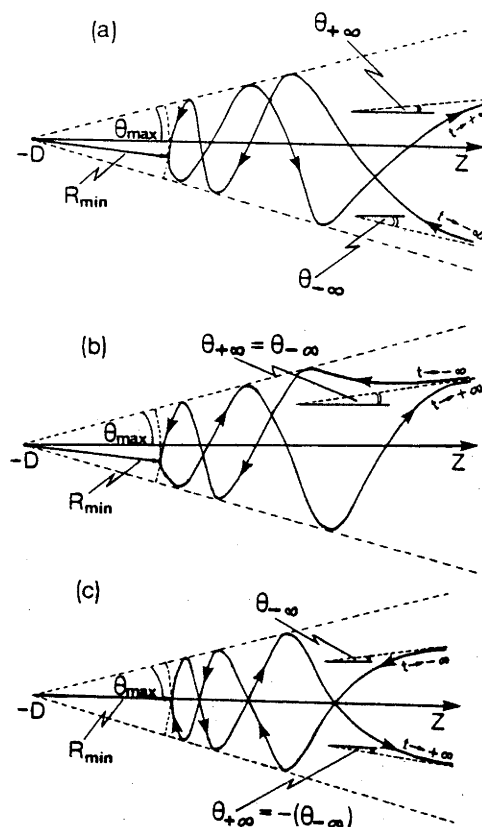


Fig. 8. Semischematic plots of meridional ray paths. A general path with seven axis crossings is shown in (a) while (b) and (c) show the special cases of perfect reflection and a symmetric path.

A similar analysis can be carried out for non-meridional rays by starting with Eq. (33).

2.10 CONCLUSION

We have shown that for graded media as in Eq. (1) there are two ray path constants, w and κ , and for the tapered quadratic-index case a complete explicit formula follows for the small-angle ray paths. The deduced behavior - concentration of rays into a small angular region, oscillations and reflections out of the taper region - will also occur for $G(\theta)$ given by any function similar to θ^2 .

The simplicity of these taper results leads to some elementary conclusions regarding the concentrator problem: designing a system to bring a given class of rays down to a given surface. If we consider a

detector surface defined by $R = R_d$, $\theta < \theta_d$, for example, then rays will be brought down onto this surface provided $R_{\min} < R_d$ and $\theta_{\max} < \theta_d$. Using Eq. (22) we see that this requires

$$\kappa < n_o^2 R_d^2 \quad (46a)$$

and since from Eq. (34) $\theta_{\max}^2 < 1/\gamma^2 = \kappa/2\Delta n_o^2$ (the equality holding for meridional rays), we also require

$$\kappa < 2\Delta n_o^2 \theta_d^2 \quad (46b)$$

Thus the rays compressed down onto the detector surface may be identified using their path constant κ . The condition (46b) is sufficient, but may not always be necessary, since when $R = R_d$ a ray may still have $\theta < \theta_d$ even though its path has $\theta_{\max} > \theta_d$.

Finally note that tapering media may be modelled with other functions $G(\theta)$, and this is discussed in later chapters. In the next chapter, we give the wave theory for the general class of GRIN media, and a detailed study of guided-wave propagation in the tapered quadratic-index medium.

2.11 REFERENCES

1. R.K. Luneburg, "Mathematical Theory of Optics," (U. California Press, Berkeley, 1964.)
2. S. Cornbleet, "Microwave and Optical Ray Geometry," (Wiley, New York, 1984).

CHAPTER 3

PROPAGATION IN A TAPERED GRADED-INDEX MEDIUM II: WAVE THEORY

3.1 INTRODUCTION

In chapter 2 we presented an exact ray analysis of a whole class of GRIN media, and in particular gave a detailed study of ray propagation in a tapered quadratic-index medium. In this chapter we give the corresponding wave theory.

The emphasis here is on obtaining a detailed description of the behavior of waves in a tapered quadratic-index medium. This is clearly very useful with regard to the various applications outlined in section 1.4.1 of chapter 1 (focusing elements, image magnifiers and reducers, light concentrators).

To ensure that the interesting propagation characteristics are not obscured by mathematical complexities, we first examine the case of tapering media in 2D. (The theory is extended to 3D media later in the chapter.) We show that those guided-wave solutions that are tightly concentrated about the axis of the taper are given, very accurately, by remarkably simple expressions which very clearly illustrate the influence of the tapering. Links with the ray theory of the last chapter are stressed.

In section 3.2 we introduce a general class of planar GRIN media (the 2D analog of that studied in chapter 2) and describe the tapered quadratic-index slab. The wave equation is solved exactly in section 3.3, and the tightly bound taper fields are obtained. The physical

characteristics of these fields are described in section 3.4, and their application discussed in section 3.5. The corresponding results for the 3D case are given in section 3.6. Some concluding remarks form section 3.7.

3.2. MODEL OF 2D TAPER

We consider the class of graded-index media modelled by the following refractive index distribution:

$$n^2 = n_0^2 \{ 1 - 2\Delta G(\theta) / R^2 \} \quad . \quad (1)$$

Here (R, θ) are plane-polar coordinates (see Fig. 1), $G(\theta)$ is an arbitrary 2π -periodic function of θ , and 2Δ is a constant with the dimensions length-squared. Equation (1) above is a 2D index distribution which corresponds to the index in the meridional plane of the general 3D GRIN medium of chapter 2. Introducing rectangular coordinates x, z as depicted in Fig. 1, we may picture the index distribution as extending unchanged in the y direction, thus giving a slab geometry.

As noted in chapter 2, a large number of fundamentally different graded-index media can be modelled by Eqn. (1), depending on the choice of $G(\theta)$, but of most interest to us here is that giving a tapered quadratic index medium:

$$G(\theta) = \sin^2\theta / \cos^4\theta \quad . \quad (2)$$

Converting to rectangular coordinates through the transformation (see Fig. 1):

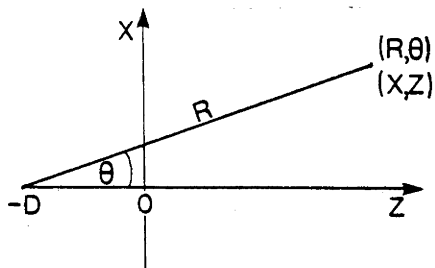


Fig. 1. Coordinate systems. The origin of the $R - \theta$ plane-polar coordinate system is located at the point $x = 0, z = -D$ of the Cartesian coordinate system.

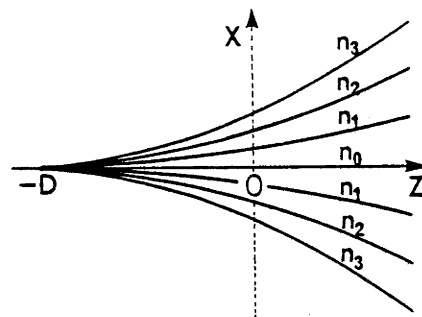


Fig. 2. Constant refractive-index contours for the graded-index taper defined by Eq. (4). Index values are ordered $n_0 > n_1 > n_2 > n_3$.

$$x = R \sin \theta$$

$$z = R \cos \theta - D$$

(3)

(D is a constant which we choose to be positive), we can rewrite Eq. (1) as

$$n^2 = n_0^2 \left\{ 1 - 2\Delta x^2 / (D+z)^4 \right\} \quad (4)$$

which describes a parabolically tapered quadratic-index slab (see section 2.2). For later purposes, we re-write Eq. (4) in the form-

$$n^2 = n_0^2 \left\{ 1 - 2\delta x^2 / \rho^2(z) \right\} \quad (5)$$

where

$$\rho(z) = \rho_0 (1 + z/D)^2 \quad (6)$$

is the characteristic radius of the grading in the transverse direction and

$$2\delta = 2\Delta \rho_0^2 / D^4 \quad (7)$$

is a dimensionless parameter. Some constant refractive index contours are reproduced in Fig. 2.

It was also pointed out in chapter 2 that the index distribution described by Eqn. (5) becomes unphysical for large values of $\sqrt{2\delta} |x| / \rho(z)$. However, this will be of no concern provided we consider only those fields which are negligible when $\sqrt{2\delta} |x| > \rho(z)$.

As discussed in section 1.5 of the introductory chapter, the electromagnetic field may be constructed (to a good approximation) from solutions to the scalar wave equation, under the weak guidance assumption¹. Assuming an implicit time dependence $\exp(-i\omega t)$ where $\omega = ck = 2\pi c/\lambda$ (λ denotes the free-space wavelength) we are lead to the Helmholtz equation:

$$(\nabla^2 + k^2 n^2) \psi = 0 \quad (8)$$

3.3 EXACT SOLUTION OF THE HELMHOLTZ EQUATION

Exact solutions for the Helmholtz equation (Eqn. (8)) can be found for all media in the form of Eqn. (1), by the method of separation of variables. Writing

$$\psi(R, \theta) = A(R) B(\theta) \quad (9)$$

we find that the R-dependence is a linear-combination of Bessel functions²:

$$A(R) = \begin{cases} J_{\mu} (kn_o R) \\ Y_{\mu} (kn_o R) \end{cases} \quad (10)$$

and the θ -dependence is given by the solution of

$$d^2B/d\theta^2 + (\mu^2 - \alpha^2 G(\theta)) B(\theta) = 0 \quad (11)$$

Here,

$$\alpha = kn_o \sqrt{2\Delta} \quad (12)$$

and μ^2 is the separation constant.

Regardless of the specific form of $G(\theta)$, the R -dependence of the field is always given by Eqn. (10). For the graded-index taper, $G(\theta)$ is given by Eqn. (2) and we find the θ -dependence of the field solutions by solving:

$$d^2B/d\theta^2 + (\mu^2 - \alpha^2 \sin^2\theta/\cos^4\theta) B(\theta) = 0 \quad (13)$$

The solution of Eqn. (13) also determines the admissible values of the separation constant, μ^2 .

Although the solution of Eqn. (13) cannot be readily expressed in terms of commonly used functions, it has a particularly simple form for small θ . This is discussed in the next section.

3.3.1 SMALL-ANGLE APPROXIMATION

To obtain the solution of Eqn. (13) for small θ , we make the approximation

$$G(\theta) \approx \theta^2 \quad (14)$$

to second order in θ , and so Eqn. (13) becomes

$$d^2B/d\theta^2 + (\mu^2 - \alpha^2\theta^2) B(\theta) = 0. \quad (15)$$

If the solution of Eqn. (15) leads to fields which are tightly concentrated about the axis of the taper, and negligible for larger values of $|\theta|$, then solving Eqn. (15) in place of Eqn. (13) will most certainly be an excellent approximation. Consequently, we seek solutions of Eqn. (15) which tend to zero for large $|\theta|$.

This restriction on the physically acceptable solutions leads to the condition³

$$\mu^2 = \alpha(2n + 1), \quad n = 0, 1, 2, \dots \quad (16)$$

and to the corresponding solutions

$$B_n(\theta) = \exp(-\alpha\theta^2/2) H_n(\sqrt{\alpha}\theta) \quad (17)$$

where the H_n functions are standard Hermite Polynomials².

We thus have a set of basis fields which can be used to describe the propagation of waves concentrated about the axis of the taper. Writing them out explicitly,

$$\psi_n(R, \theta) = \exp(-\alpha\theta^2/2) H_n(\sqrt{\alpha}\theta) \begin{cases} J_{\sqrt{\alpha(2n+1)}}(kn_o R) \\ Y_{\sqrt{\alpha(2n+1)}}(kn_o R) \end{cases} \quad (18)$$

(It should be mentioned that the Hankel functions² $H_{\mu}^{(1)}(kn_o R)$ and $H_{\mu}^{(2)}(kn_o R)$ (where $\mu = \sqrt{\alpha(2n+1)}$) could be used in place of the J and Y Bessel functions, since both sets obviously provide two linearly-independent solutions of the R-equation.)

Clearly, the basis fields have an unchanging field shape over any radial arc (i.e. at any fixed R), and in this sense are a generalization of the mode concept of translationally-invariant media (this is discussed again in section 3.4.2 when we consider the basis fields from the point of view of transverse resonance).

The actual combination of the Bessel functions to be used depends upon the boundary conditions (i.e. the region of the taper in which propagation occurs). This follows from the requirement that the basis fields must be able to exist in isolation of each other (so that they can, at least in principle, be excited individually). This point will become clearer when several applications of the basis fields are discussed in section 3.5.

3.3.2 COMPARISON WITH UNTAPERED MEDIUM

Because of their simplicity, the basis fields demonstrate the effect that the tapering has on wave propagation in the quadratic-index medium with remarkable clarity. We compare the basis fields (Eqn. (18)) with the modes of a translationally invariant quadratic-index medium:

$$n^2 = n_o^2 \{1 - 2\delta x^2/\rho_o^2\} \quad (19)$$

(Here ρ_o is the constant characteristic radius of the medium - compare with Eqns. (5) and (6).) The propagating Hermite-Gaussian modes of such a medium are well-known⁴:

$$\psi_{n(\text{inf.par.})} = \exp(-qx^2/2\rho_o) H_n[(q/\rho_o)^{1/2} x] \exp(\pm i\beta_n z) \quad (20)$$

$$n = 0, 1, 2, \dots$$

where

$$\beta_n = \{k^2 n_o^2 - (q/\rho_o)(2n + 1)\}^{1/2} \quad (21)$$

are the propagation constants, and the parameter q is defined by

$$q = kn_o \sqrt{2\delta} \quad (22)$$

We make the comparison with the basis fields (Eqn. (18)) by converting the latter to rectangular coordinates. Noting that within the small-angle approximation

$$R \approx D + z \quad (23)$$

$$\theta \approx x/(D + z)$$

then the basis fields in Eqn. (18) may be written:

$$\psi_n(x, z) = \exp[-qx^2/2\rho(z)] H_n[(q/\rho(z))^{1/2} x] \begin{cases} J \\ Y \end{cases} \frac{1}{D\sqrt{q(2n+1)/\rho_o}} (kn_o(D+z)) \quad (24)$$

Comparing Eqns. (20) and (24) reveals that in any cross-section of the taper the field distribution is identical to that within a straight quadratic-index medium which matches the taper cross-section at that particular value of z . (This, incidentally, illustrates the power of local mode approaches to wave propagation in tapered media). Only the propagation characteristics of the two sets of fields differ: in the untapered case propagation is described by complex exponentials, whereas for tapered media it is described by some combination of Bessel functions.

We can also check the limiting behaviour of the basis fields by letting the degree of tapering become negligible. This is formally carried out by finding the asymptotic form of Eqn. (24) for large D (provided we constrain z to lie within fixed limits i.e. we consider a finite section of the taper). From the definition of the taper profile (Eqns. (5) and (6)) it is clear that this limit reduces the medium to the translationally invariant profile Eqn. (19).

In the Appendix we show that the basis fields with radial dependence given by the Hankel functions, do indeed become the forward and backward propagating modes of an untapered quadratic-index medium in the large D limit. Thus the basis fields do display the correct limiting behavior.

3.4 PROPERTIES OF THE BASIS FIELDS

As mentioned previously, the basis fields are tightly concentrated about the axis of the taper. To illustrate this, in Fig. 3 we have plotted $B_n(\theta)$ (Eqn. (17)) for $n = 0$ and $n = 5$ and with typical parameter values. In general, $B_n(\theta)$ has an oscillatory part centred about the origin (with n zeros) and has an exponential drop-off at larger

values of $|\theta|$ (the "wings" of the distribution). Thus the field may be considered confined to the region

$$-\tilde{\theta}_n < \theta < \tilde{\theta}_n \quad (25)$$

where $\tilde{\theta}_n$ is some value of θ in the wings of the distribution where the decaying exponential behaviour has become dominant. We can quantify this by choosing $\tilde{\theta}_n$ to be the value of θ where B_n changes from concave downwards to concave upwards (i.e. $d^2B_n / d\theta^2 = 0$) and from Eqn. (15):

$$\tilde{\theta}_n = \{(2n + 1) / \alpha\}^{1/2} \quad (26)$$

Thus, the field is confined to the region inside a wedge of half-angle $\tilde{\theta}_n$, and from Eqn. (26) it is clear that basis fields with larger values of n are spread out more than those with smaller values. This leads us to a simple condition for the validity of the basis field expressions:

$$\{(2n + 1) / \alpha\}^{1/2} \ll 1 \quad (27)$$

When expression (27) is not satisfied, serious errors may be incurred by solving Eqn. (15) in place of Eqn. (13).

The properties of the R-dependence of the field will be discussed later when several specific examples are given (section 3.5).

3.4.1 COMPARISON WITH GEOMETRICAL OPTICS

The ray solution for the slab-geometry taper corresponds to the meridional rays of the 3D taper studied in chapter 2. In this section we compare the ray analysis with the characteristic behavior of the basis fields. A valuable physical interpretation of several results from the wave theory may be obtained through such ray-mode links.

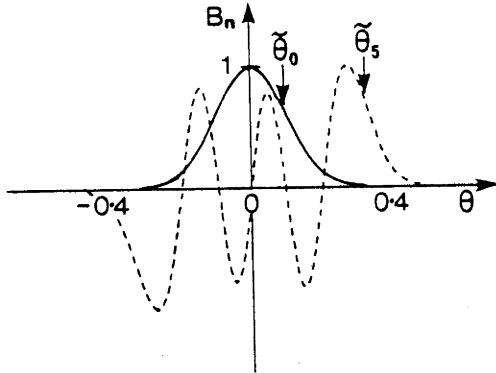


Fig. 3. Plots of $B_n(\theta)$ for $n = 0$ (solid line) and $n = 5$ (broken line) and with $\alpha = 100$. (Note that the curves have been normalized to give a maximum value of unity.) Also shown are $\tilde{\theta}_0$ and $\tilde{\theta}_5$ where the second derivative disappears.

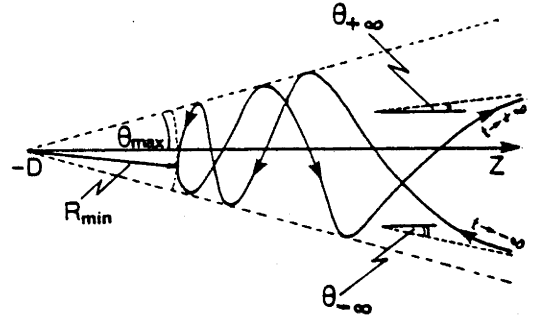


Fig. 4. Schematic drawing of a typical ray trajectory inside the graded-index taper. Ray caustics are shown at $R = R_{\min}$ and $\theta = \pm\theta_{\max}$, confining the ray to the region inside a truncated wedge. Far from the taper apex the ray tends asymptotically to a straight line, here denoted $\theta_{-\infty}$ (as the ray comes in) and $\theta_{+\infty}$ (as the ray leaves).

Recall from chapter 2 that an invariant for each ray is given by (see Eq. (2.20))

$$\kappa = R^2(t) \{n_0^2 - \dot{R}^2(t)\} = R^2(0) \{n_0^2 - \dot{R}^2(0)\} \quad (28)$$

where t is a parameter defined by $dt = dS/n$ (S denotes length measured along the ray path) and the dot convention is used to denote a t -derivative. The θ -variable is obtained from the equation:

$$R^4 \dot{\theta}^2 = \kappa - 2\Delta n_0^2 G(\theta) \quad (29)$$

(Eqn. (2.14) with skew invariant $w = 0$). A typical ray propagating into the taper is illustrated in Fig. 4. Note that rays are confined to the region:

$$-\theta_{\max} < \theta < \theta_{\max} \quad (30)$$

$$R_{\min} < R \quad (31)$$

as shown in the figure. Here, $\theta_{\max}$ and $R_{\min}$ are determined by the initial launching conditions of the ray. It is found that $R_{\min}$ is exactly given by (see Eqn. (2.22)):

$$R_{\min} = \sqrt{\kappa} / n_o \quad (32)$$

while $\theta_{\max}$ is given, in the small-angle approximation, by (see Eqn. (2.39)):

$$\theta_{\max} = \sqrt{\kappa} / n_o \sqrt{2\Delta}. \quad (33)$$

All rays with the same value ray invariant, κ , are confined to exactly the same region - the truncated wedge defined by Eqns. (30)-(33). The family of such rays are said to be confined by caustic surfaces.

We immediately note the agreement between the ray and field analyses concerning the confinement of the light to the inside of a wedge. Comparing $\theta_{\max}$ of the ray analysis (Eqn. (33)) with $\tilde{\theta}_n$ of the wave theory (Eqn. (26)) allows us to associate a basis field ψ_n with the family of rays with ray-invariant κ . Setting $\theta_{\max} = \tilde{\theta}_n$ we obtain

$$\kappa = \alpha (2n + 1) / k^2 \quad (34)$$

The above relation gives the criterion for a family of rays to correspond to a basis field. This result is developed more formally in the next section.

Finally we point out that there are other striking agreements between the ray and field pictures which will become evident when we look at applications of the basis fields, in section 3.5.

3.4.2 BASIS FIELD AS A TRANSVERSE RESONANCE

In this section we provide a simple and intuitive explanation of the ray-field relationship expressed in Eqn. (34). This will be a straightforward generalization of the "transverse resonance" concept that has been successfully applied to translationally invariant waveguides⁴.

We begin by noting that when we associate the ray description of light propagation with the wave description, we must introduce phase with the ray trajectories. Rays become associated with "local plane-waves" with propagation vectors directed along the ray paths¹.

If (R, θ) are the plane-polar coordinates of a point on the trajectory, then the local plane-wave propagation vector is

$$\tilde{k}_{loc}(R, \theta) = k n(R, \theta) \left\{ \frac{\dot{R}}{\sqrt{(\dot{R})^2 + R^2 \dot{\theta}^2}} \hat{e}_R + \frac{R \dot{\theta}}{\sqrt{(\dot{R})^2 + R^2 \dot{\theta}^2}} \hat{e}_\theta \right\} \quad (35)$$

(where the dot convention is used as in section 3.4.1 and $\hat{e}_R, \hat{e}_\theta$ are plane-polar unit-vectors in the radial and transverse-radial directions).

The basis fields (Eqn. (18)) are characterized by an unchanging field distribution over any radial arc. This indicates that the $\hat{e}_\theta$ component of the local plane-wave vector must satisfy a self-consistency condition in order to describe a basis field. In Fig. 5 we show a typical ray path over a single period ($S = S_0 \rightarrow S = S_2$ where S denotes path length along the ray trajectory). Clearly, in order to describe a basis

field, the phase change of the $\hat{e}_\theta$ component due to the optical path length traversed, added to the phase change suffered due to reflection at the caustics (at S_0 and S_1) must be equal to an integer multiple of 2π . This is the transverse resonance condition. Noting that reflection at a caustic⁵ introduces a phase change of $-\pi/2$, we have:

$$\Delta\Phi + 2(-\pi/2) = 2\pi n \quad (36)$$

(n an integer)

where

$$\Delta\Phi = \int_{\text{ray path}}^{S_0 \rightarrow S_2} \{k n(R, \theta) R \dot{\theta} / [R^2 + R^2 \dot{\theta}^2]^{1/2}\} R d\theta \quad (37)$$

Using Eqns (28) and (29) we can simplify Eqn. (37) to

$$\Delta\Phi = 2\alpha \int_{-\theta_{\max}}^{+\theta_{\max}} \{G(\theta_{\max}) - G(\theta)\}^{1/2} d\theta \quad (38)$$

Within the small-angle approximation we can replace $G(\theta)$ and $G(\theta_{\max})$ with θ^2 and $\theta_{\max}^2$ respectively. A simple integration then yields

$$\Delta\Phi = \pi k^2 \kappa / \alpha \quad (39)$$

Substitution of Eqn. (39) into the transverse resonance condition (Eqn. (36)) leads to the ray-basis field relationship, Eqn. (34).

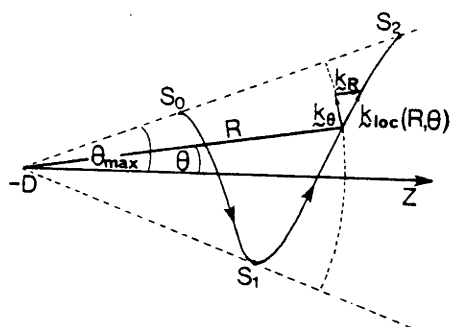


Fig. 5. Single period of a typical ray path is shown from S_0 to S_2 . Also illustrated is the local plane-wave vector $k_{loc}(R, \theta)$ at a point (R, θ) on the trajectory and its radial component k_R and transverse-radial component k_θ .

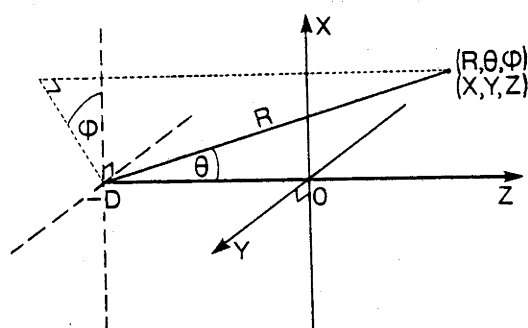


Fig. 7. Coordinate systems. The origin of the $R - \theta - \phi$ spherical-polar coordinate system is located at the point $x = 0, y = 0, z = -D$ of the Cartesian system. We follow the usual convention: $0 \leq R < \infty$; $0 \leq \theta \leq \pi$; $0 \leq \phi < 2\pi$.

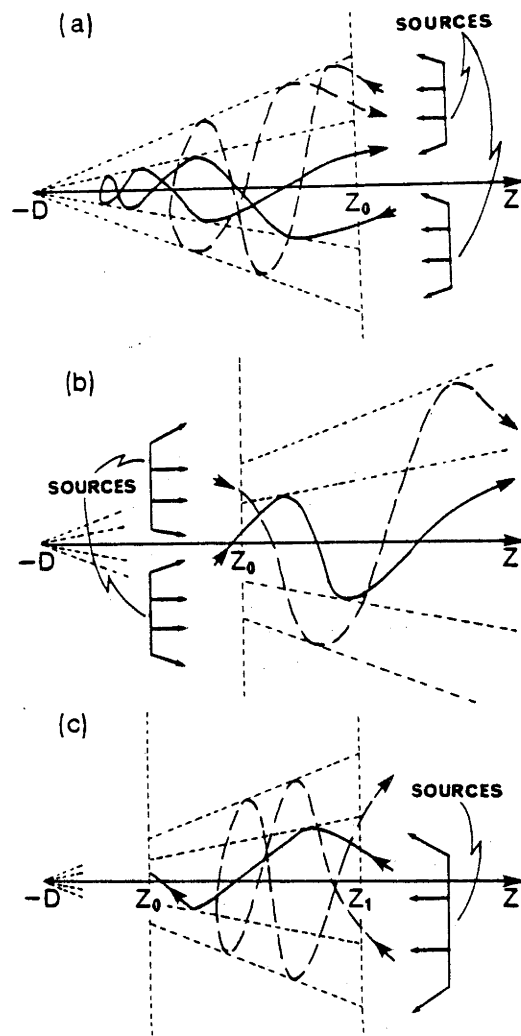


Fig. 6. Schematic illustrations of the examples discussed in Sec. V.A (a), V.B (b), and V.C (c). Also shown are some typical ray trajectories encountered in these cases.

3.5 EXAMPLES OF APPLICATIONS OF THE BASIS FIELDS

In this section we present three simple examples illustrating the application of the basis fields.

3.5.1 FIELDS LAUNCHED DOWN THE COMPLETE TAPER

As a first example, suppose we look at propagation in the complete graded-index taper (i.e. we assume that the taper is not truncated anywhere so that it narrows to a point). With all sources located in the $z > z_0$ region, we are interested in determining the field in the $z < z_0$ region. This example is illustrated schematically in Fig. 6(a). (The

reader is warned that there are some physical difficulties associated with the mathematical model described below. They will be discussed at the conclusion of this section.)

In this case, because the taper extends to $R = 0$, the Y_μ -part of the radial-dependence (Eqn. (10)) must be rejected (since these solutions become unbounded for small R). Hence, the basis fields take the explicit form:

$$\psi_n(R, \theta) = \exp(-\alpha\theta^2/2) H_n(\sqrt{\alpha}\theta) J_{\frac{\mu}{\sqrt{\alpha(2n+1)}}}(kn_o R). \quad (40)$$

These fields are clearly standing-wave type solutions; the net time-averaged power passing through any given surface with fixed R is zero. Furthermore, from the general behaviour of Bessel functions we can infer a further important characteristic of these fields. Firstly we note the well-known fact that the Bessel function $J_\mu(x)$ ($\mu \gg 1$) has rather different asymptotic behaviour in the regions $x \ll \mu$ and $x \gg \mu$. In the region $x \ll \mu$ it behaves like a decaying exponential, whereas it has oscillatory behaviour when $x \gg \mu$. This leads us to conclude that the basis fields must be considered confined to the region

$$R > \tilde{R}_n \quad (41)$$

where $\tilde{R}_n$ is some value of R beyond which the decaying exponential behaviour becomes dominant (i.e. the field becomes evanescent). Taking equality between order and argument to denote this value⁶, we have

$$\tilde{R}_n = \{\alpha(2n+1)\}^{1/2} / kn_o \quad (42)$$

We shall refer to the surfaces $R = \tilde{R}_n$ and $\theta = \tilde{\theta}_n$ as wave caustics because they can be associated with the ray caustics $R = R_{\min}$ and $\theta = \theta_{\max}$ respectively. (Strictly speaking, the wave caustic occurs where the field is a maximum, not at the point of inflection. In practice, however, the difference is often negligible.)

There is a very simple, physical explanation of this result. Writing

$$J_{\mu}(kn_0 R) = (1/2) \{H_{\mu}^{(2)}(kn_0 R) + H_{\mu}^{(1)}(kn_0 R)\} \quad (43)$$

allows us to interpret the basis field as a backward-propagating field (with radial dependence $H_{\mu}^{(2)}(kn_0 R)$) travelling into the taper and reflecting off the caustic at $\tilde{R}_n$, thereby converting it into a forward-propagating field (with radial-dependence $H_{\mu}^{(1)}(kn_0 R)$). The backward and forward-propagating fields interfere with each other to create the standing wave field. The location of the caustic (Eqn. (42)) corresponds physically to the cut-off of the associated mode of the local un-tapered quadratic-index medium with cross-section matching the tapered medium at z (i.e. ρ_0 replaced by $\rho(z)$ in Eqn. (19)). From Eqn. (21) we see that an un-tapered quadratic-index medium with characteristic radius $\rho(z)$ supports modes with propagation constants

$$\beta_n = \{k^2 n_0^2 - (q/\rho(z))(2n+1)\}^{1/2} \quad (44)$$

When the characteristic radius is too small, the propagating mode is cut-off and becomes evanescent (i.e. β_n becomes imaginary). From Eqns (44) and (23) and the definitions of the various parameters, we find that this occurs at $\tilde{R}_n$ given by Eqn. (42). Thus, the caustic occurs where the taper is too narrow to support the n th propagating mode.

Thus, for this first example, we have shown that the basis fields are confined to the region inside a truncated wedge ($-\tilde{\theta}_n < \theta < \tilde{\theta}_n$ and $R > \tilde{R}_n$). This is clearly in agreement with the ray theory: all rays launched into the $z < z_0$ region are also confined to a truncated wedge, as illustrated in Fig. 4. They oscillate down the taper, reflect off the caustic, and then oscillate out again. Comparing the field caustic $\tilde{R}_n$ (Eqn. (42)) with the ray caustic $R_{\min}$ (Eqn.(32)) gives exactly the ray-basis field relationship of Eqn. (34).

A general field launched down the taper is expressible as a superposition of the basis fields (provided that this incident field is sufficiently concentrated about the taper axis). Because of the finite number of physically meaningful basis fields (see Eqn. (27)) this is clearly not an exact representation, but will be a very good approximation for many situations of practical interest. Remembering that on any particular plane each basis field has both forward and backward-propagating components (Eqn.(43)) we must match the incident field to the backward-propagating ($-z$ direction) component and use the orthogonality of Hermite polynomials² to obtain the amplitude of each basis field appearing in the superposition. The field reflected out of the taper can then be found.

Unfortunately the mathematical model we have described is inadequate at two small regions near the caustic surface: the index distribution is unphysical (less than unity) in a small region near each cusp of the caustic surface (i.e. near the points $(\tilde{R}_n, \pm\tilde{\theta}_n)$). Therefore some errors will be incurred when we use this model to represent a real physical graded-index taper. Nevertheless the model is still useful in illustrating the physics behind the reflection of waves out of a taper.

3.5.2 FIELDS LAUNCHED UP THE COMPLETE TAPER

For the second example, we once again consider propagation in the complete graded-index taper, but this time all sources are located in the $z < z_0$ region and we are interested in the field in the region $z > z_0$. We illustrate this schematically in Fig. 6(b).

It is more convenient, in this case, to consider the R -dependence of the basis fields to be a linear combination of the $H_{\mu}^{(1)}(kn_0 R)$ and $H_{\mu}^{(2)}(kn_0 R)$ Hankel functions (as discussed in section 3.3.1). This is because, from a consideration of the asymptotic form of these functions² for large R :

$$H_{\mu}^{(1),(2)}(kn_0 R) \sim (2/\pi kn_0 R)^{1/2} \exp\{\pm i[kn_0 R - \mu(\pi/2) - \pi/4]\} \quad (45)$$

(where the $+/-$ refers to $H_{\mu}^{(1)}$ and $H_{\mu}^{(2)}$ respectively) it is clear that we must reject the $H_{\mu}^{(2)}$ solution because there can be no backward-propagating (i.e. in the $-z$ direction) field at infinity. (As noted in section 2 we are assuming a time dependence $\exp(-i\omega t)$). Hence, in this example the basis fields take the form:

$$\psi_n(R, \theta) = \exp(-\alpha\theta^2/2) H_n(\sqrt{\alpha}\theta) \frac{H_{\mu}^{(1)}(kn_0 R)}{\sqrt{\alpha(2n+1)}} \quad (46)$$

These fields represent progressive waves, and in the limit of negligible tapering become the forward-propagating modes of the un-tapered quadratic-index medium (see section 3.3.2, and the appendix).

Once again, it should be mentioned that a general field launched up the taper can be represented (approximately) by a superposition of the basis fields, Eqn. (46). However, in this case the problem is simplified by the absence of any back-scattered field components so that the incident field (on the plane $z = z_0$) is expanded directly in terms of the functions in Eqn. (46) evaluated on this plane.

3.5.3 FIELDS LAUNCHED IN A SECTION OF THE TAPER

The problem of most practical interest concerns wave propagation inside a section of the taper (say between the planes $z = z_0$ and $z = z_1 > z_0$, as illustrated schematically in Fig. 6(c)), and assuming that all sources lie outside of this region.

Within the restrictions discussed in the previous examples, a general field incident on either end-face ($z = z_0$ or $z = z_1$) will create a field inside the taper section which is expressible as a superposition of the basis fields:

$$E = \sum_n a_n \exp(-\alpha\theta^2/2) H_n(\sqrt{\alpha}\theta) \left\{ J_{\frac{(kn_0 R)}{\sqrt{\alpha(2n+1)}}} + b_n Y_{\frac{(kn_0 R)}{\sqrt{\alpha(2n+1)}}} \right\} \quad (47)$$

where the constants a_n and b_n are complex amplitudes determined from boundary conditions.

Some of the added complications in dealing with a finite section of the taper are made clearer by looking at the problem from the viewpoint of geometrical optics. Considering the situation shown in Fig. 6(c) with the sources in the region $z > z_1$, we see that some of the launched rays will pass right through the taper (when the ray caustic lies outside of the taper section) while others will be turned around and exit the same way

they entered (ray caustic inside the taper). Thus, even ignoring reflections from the interfaces, the problem is clearly more complicated than in the previous two examples with the complete taper.

3.6 PROPAGATION INSIDE A TAPER OF CIRCULAR GEOMETRY

We will now very briefly analyse the situation occurring inside a graded-index taper of circular geometry. Results are very much the same as for the planar taper, but with the added complication of a further dimension.

As in the planar case, we start by defining a general class of graded-index media:

$$n^2 = n_0^2 \{1 - 2\Delta G(\theta) / R^2\} \quad (48)$$

but this time (R, θ, ϕ) are spherical-polar coordinates (see Fig. 7). The description of this general class of media was given in section 2.2 of the previous chapter. Here we give the wave theory for this class of GRIN media, and obtain simple expressions for the tightly bound fields of a tapered quadratic-index medium of circular geometry, by choosing

$$G(\theta) = \sin^2\theta / \cos^4\theta. \quad (49)$$

(See the discussion in section 2.2.) Again, we assume weak guidance¹, an implicit time dependence $\exp(-i\omega t)$, and seek solutions to the Helmholtz equation (Eqn. (8)).

3.6.1 EXACT SOLUTION OF THE HELMHOLTZ EQUATION

Using the method of separation of variables, we can find exact solutions for all media having an index distribution in the form of Eqn. (48). Writing

$$\psi(R, \theta, \phi) = A(R) B(\theta) C(\phi) \quad (50)$$

we find that

$$A(R) = \begin{cases} (1/\sqrt{R}) J_{\frac{(kn_0 R)}{\sqrt{\nu+(1/4)}}} \\ (1/\sqrt{R}) Y_{\frac{(kn_0 R)}{\sqrt{\nu+(1/4)}}} \end{cases} \quad (51)$$

$$C(\phi) = \begin{cases} \cos m\phi \\ \sin m\phi \end{cases}, \quad m = 0, 1, 2, \dots \quad (52)$$

and $B(\theta)$ is found by solving

$$\begin{aligned} \sin^2 \theta \, d^2 B / d\theta^2 + \sin \theta \cos \theta \, dB / d\theta \\ + \{ \nu \sin^2 \theta - \alpha^2 \sin^2 \theta G(\theta) - m^2 \} B(\theta) = 0 \end{aligned} \quad (53)$$

The solution of Eqn. (53) also determines the admissible values of the separation constant, ν .

For the graded-index taper $G(\theta)$ is given by Eqn. (49). However, as in the planar case, the solution of Eqn. (53) cannot be readily expressed in terms of commonly used functions. Fortunately, however,

within the small-angle approximation the solutions take on a particularly simple form.

3.6.2 THE SMALL-ANGLE APPROXIMATION

Under the small-angle approximation we have $G(\theta) \approx \theta^2$ (see Eqn.(14)), and we likewise expand the coefficients appearing in Eqn. (53) to second order in θ . This leads to the simplified equation:

$$\theta^2 \frac{d^2 B}{d\theta^2} + \theta \frac{dB}{d\theta} + \{v\theta^2 - \alpha^2 \theta^4 - m^2\} B(\theta) = 0 \quad (54)$$

The solution of this equation will provide a good description of those field solutions which are concentrated about the axis of the taper. We are thus lead to seek solutions of Eqn. (54) which tend to zero for large θ . It is found that the only solutions which give fields bounded on the taper axis while satisfying the above restriction, are of the form:

$$B_{n,m}(\theta) = \exp(-\alpha\theta^2/2) \theta^m L_n^{(m)}(\alpha\theta^2) \quad (55)$$

$$m, n = 0, 1, 2, \dots$$

where $L_n^{(m)}$ are Associated Laguerre Polynomials as defined in Abramowitz and Stegun². The separation constant takes on the discrete values:

$$v = 2\alpha (m + 2n + 1) . \quad (56)$$

Thus we have found a set of basis fields describing wave propagation close to the axis of the graded-index taper of circular geometry. Writing them out explicitly:

$$\psi_{n,m}^{o,e} = R^{-1/2} \theta^m \exp(-\alpha\theta^2/2) L_n^{(m)}(\alpha\theta^2) \begin{cases} \sin(m\phi) & J_\zeta(kn_o R) \\ \cos(m\phi) & Y_\zeta(kn_o R) \end{cases}$$

$$m, n = 0, 1, 2, \dots$$

(57)

where

$$\zeta = (1/2) \{1 + 8\alpha(m+2n+1)\}^{1/2}. \quad (58)$$

Note that the superscripts "o" and "e" which appear in the symbol ψ , denote the basis fields which are odd/even with respect to ϕ . Thus, for $\psi_{n,m}^o$ the ϕ -dependence is given by $\sin m\phi$, and for $\psi_{n,m}^e$ by $\cos m\phi$. (For $m \neq 0$ these are degenerate of course: $\psi_{n,m}^o$ and $\psi_{n,m}^e$ have identical field distributions, the only difference being in a rotation by 90° about the axis of the taper.

The behavior of these basis fields is completely analogous to the planar case considered in earlier sections.

3.7 CONCLUSION

The basis fields derived in this chapter are very accurate descriptions of those field solutions which are tightly concentrated about the axis of the taper. Due to their remarkable simplicity, the effects associated with tapering a quadratic-index medium are revealed with great clarity. Of particular interest here has been the compression and subsequent reflection of waves propagating into the taper.

The basis fields may also be applied in a variety of studies - for example, accurate analysis of focusing and imaging in tapered GRIN media. In addition, the results presented in this chapter make available a simple theoretical model which may be used to test the accuracy of approximation methods which are used extensively in the literature (for example, coupled modes and adiabatic methods, slowly varying envelope approaches).

3.8 APPENDIX: LIMITING BEHAVIOR OF THE BASIS FIELDS

Here we outline the procedure for examining the behaviour of the basis fields in the limit of negligible tapering.

From the definition of the index distribution (Eqns. (5) - (6)) we keep 2δ and ρ_0 fixed, and allow $\rho(z)$ to depend parametrically on D . Clearly, the degree of tapering is determined by D . If we consider a given section of the taper (so that z is constrained to lie within fixed limits) and let $D \rightarrow \infty$, then the tapering disappears and we are left with a section of translationally invariant, infinite quadratic-index medium (Eqn. (19)). We now show that the basis fields (Eqn. (24)) exhibit the correct limiting behaviour.

Firstly, we note that we can write the basis fields (Eqn. (24)) in the form

$$\psi_n = \exp(-q x^2 / 2 \rho(z)) H_n \left[(q/\rho(z))^{1/2} x \right] \begin{cases} J_\mu(\mu \sec \delta) \\ Y_\mu(\mu \sec \delta) \end{cases} \quad (A1)$$

where

$$\mu = D \{ q(2n+1) / \rho_0 \}^{1/2} \quad (A2)$$

and

$$\sec\delta = kn_o \{ \rho_o / q(2n+1) \}^{1/2} (1 + z/D). \quad (A3)$$

We will assume for ease of calculation that

$$kn_o \{ \rho_o / q(2n+1) \}^{1/2} \gg 1 \quad (A4)$$

(and hence $\sec\delta \gg 1$).

The asymptotic form of Eqn. (A1) for large D follows once we know the asymptotic form of $J_\mu(\mu \sec\delta)$ and $Y_\mu(\mu \sec\delta)$ for large μ , (since from Eqns. (A2) and (A3) it is clear that taking the asymptotic limit for large D is equivalent to taking the asymptotic limit for large μ with $\sec\delta$ fixed). The standard result is²

$$J_\mu(\mu \sec\delta) \sim (2/\pi\mu \tan\delta)^{1/2} \cos(\mu \tan\delta - \mu\delta - \pi/4) \quad (A5)$$

and similarly for $Y_\mu(\mu \sec\delta)$. Under the simplifying assumption (Eqn. (A4)) we find that to first order in z/D ,

$$\mu \tan\delta \approx C_1 + D \{ kn_o + \sqrt{2\delta}(2n+1) / 2\rho_o \} (z/D) \quad (A6)$$

$$\mu \delta \approx C_2 + D \{ \sqrt{2\delta}(2n+1) / \rho_o \} (z/D)$$

(where C_1 and C_2 are independent of z).

Thus if we neglect the z -variation of the function multiplying the cosine in Eqn. (A5) (it has a very slow variation with z compared to the rapidly oscillating cosine) then we find that

$$J_{\mu}(\mu \sec \delta) \sim A_1 \cos \{ [kn_0 - \sqrt{2\delta} (2n+1)/2\rho_0] z + \phi_1 \} \quad (A7)$$

and similarly

$$Y_{\mu}(\mu \sec \delta) \sim A_1 \sin \{ [kn_0 - \sqrt{2\delta} (2n+1)/2\rho_0] z + \phi_1 \} \quad (A8)$$

(Here A_1 and ϕ_1 are amplitude and phase terms which we take to be independent of z).

With the further observation that $\rho(z) \rightarrow \rho_0$ for large D , we obtain the required limiting expression for the basis fields (Eqn. (A1)):

$$\psi_n \sim \exp(-qx^2/2\rho_0) H_n[(q/\rho_0)^{1/2} x] \quad (A9)$$

$$\times \begin{cases} \cos \{ [kn_0 - \sqrt{2\delta} (2n+1)/2\rho_0] z + \phi_1 \} \\ \sin \{ [kn_0 - \sqrt{2\delta} (2n+1)/2\rho_0] z + \phi_1 \} \end{cases}$$

We compare expression (A9) with the modes of an untapered infinite quadratic-index medium (Eqns. (20) and (21)). Under the simplifying assumption (Eqn. (A4)) we find that the modal propagation constants (Eqn. (21)) reduce to

$$\beta_n \approx kn_0 - \sqrt{2\delta} (2n+1)/2\rho_0 \quad (A10)$$

and we conclude that the basis fields do indeed have the appropriate limiting behaviour. Using the Hankel functions $H_{\mu}^{(1)}$ and $H_{\mu}^{(2)}$ in place of the J_{μ} and Y_{μ} Bessel functions, leads to the complex exponential, propagating mode solutions of Eqn. (20).

3.9 REFERENCES

1. A.W. Snyder and J.D. Love, "Optical Waveguide Theory", (Chapman and Hall, London, 1983).
2. M. Abramowitz and I.A. Stegun (eds.), "Handbook of Mathematical Functions", (Dover, New York, 1972).
3. H. Jeffreys and B.S. Jeffreys, "Methods of Mathematical Physics", (Cambridge University Press, London, 1966).
4. H. Kogelnik, "Theory of Dielectric Waveguides", in "Integrated Optics", T. Tamir, ed. Topics in Applied Physics, vol. 7, (Springer-Verlag, Berlin, 1975).
5. J.B. Keller and S.I. Rubinow, "Asymptotic Solution of Eigenvalue Problems", Ann. Phys. 9, 24 (1960).
6. The Bessel function $J_\mu(x)$ has a point of inflection (i.e. changes from concave upwards to concave downwards) roughly where the argument equals the order ($x = \mu$). In fact, if j_μ'' denotes the exact location of the smallest positive zero of $d^2 J_\mu(x) / dx^2$, then $\sqrt{\mu(\mu-1)} < j_\mu'' < \sqrt{\mu^2-1}$. See, for example, G.N. Watson, "A Treatise on the Theory of Bessel Functions", (Cambridge University Press, London, 1966). Clearly, as μ gets larger, j_μ'' / μ moves closer and closer to unity.

CHAPTER 4

RAY PROPAGATION AND COMPRESSION IN A STRICTLY ADIABATIC TAPER

4.1 INTRODUCTION

The last two chapters dealt with propagation in a tapered quadratic-index medium. This is a useful model for elucidating many important principles underlying the behavior of light in tapering structures, but is idealized in the sense that the profile extends to infinity and only makes sense physically when $\sqrt{2\delta} |x| / \rho(z) < 1$ (see sections 2.2 and 3.2). Field solutions may be chosen which are negligible outside of this region, but as a consequence of the unphysical behavior of the index profile at infinity, there is no radiation field. (The same is true, of course, for an untapered, infinite quadratic-index medium¹.) Clearly, the effects arising from the presence of a uniform cladding are of interest.

Marcatili² has given wave theory for some GRIN tapers, with uniform cladding, which guide light without radiation (see also section 1.4.2). Here the interface plays a crucial role in the guidance. The purpose of the present chapter is to give an exact ray analysis of the simplest such taper - a linearly tapered slab with radial grading - in order to obtain a clearer, physical understanding of how lossless guidance takes place. In addition, the compression of rays inside the taper is analysed exactly, and is discussed from the viewpoint of concentrator theory.

Following this introduction, we describe the strictly adiabatic taper model in section 4.2. Section 4.3 gives the ray equations (we use the formalism developed in chapter 2) and the complete ray theory follows in section 4.4. In section 4.5 we analyse the taper using concentrator theory. Interesting links between the ray and wave theories are discussed in section 4.6. Finally, section 4.7 gives exact ray theory for the 3D taper formed by rotating the planar taper about its axis of symmetry - tunnelling rays are introduced. Section 4.8 gives concluding remarks.

4.2 TAPER MODEL

We introduce plane-polar coordinates, $R-\theta$, as illustrated in Fig. 1. The refractive index distribution is given by

$$n^2 = \begin{cases} n_o^2 \{1 + N_o^2/R^2\} , & |\theta| \leq \theta_o \\ n_o^2 & , \theta_o < |\theta| \leq \pi \end{cases} \quad (1)$$

Here N_o is a constant with dimensions of length, and n_o is the uniform refractive index of the medium surrounding the taper (the cladding). The taper itself is a wedge of half-angle θ_o , graded according to Eqn. (1) so that the refractive index increases as the taper narrows.

The refractive index inside the wedge is always greater than in the uniform region outside, and waveguiding effects are to be expected. However, the field analysis by Marcatili² gives the perhaps surprising result that a finite number of lossless bound modes exists in addition to the radiation field. The ray analysis that follows will give some insight into how this occurs.

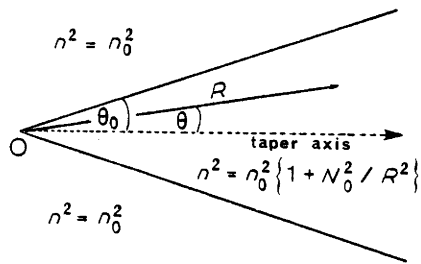


Figure 1 Co-ordinate system. Plane polar co-ordinates, $R-\theta$, are defined relative to the axis of the taper, with origin, O , at the apex.

Rewriting Eqn. (1) in the form:

$$n^2 = n_0^2 \left\{ 1 - N_0^2 G(\theta) / R^2 \right\} \quad (2)$$

where

$$G(\theta) = \begin{cases} -1, & |\theta| \leq \theta_0 \\ 0, & \theta_0 < |\theta| \leq \pi \end{cases} \quad (3)$$

it is clear that the taper belongs to the general class of graded-index media analysed in chapters 2 and 3, and that the formalism of those chapters may be applied here.

4.3 THE RAY EQUATIONS

Following chapter 2, we introduce a parameter t defined by

$$dt = dS / n \quad (4)$$

where S denotes length measured along the ray path. We describe the ray path parametrically by $R(t)$, $\theta(t)$ and use the convention of dotting a quantity to denote d/dt .

Three initial conditions required are:

$$R(t=0) = R_1 \quad (5a)$$

$$\theta(t=0) = \theta_1 \quad (5b)$$

$$\dot{R}(t=0) = \dot{R}_1 \quad (5c)$$

The physical meaning of the last initial condition was pointed out in chapter 2, where it was noted that

$$\dot{R} = n \, dR/dS = n \cos\beta \quad (6)$$

where β is the angle between the ray path and the radial direction, as illustrated in Fig. 2(a). (Also shown, in Fig. 2(b), is the complementary ray angle, $\gamma = \pi - \beta$, which is more convenient to use when the ray propagates towards the apex of the taper.) If β_1 denotes the initial ray angle, then

$$\dot{R}_1 = n(R_1, \theta_1) \cos\beta_1 \quad (7)$$

The ray equations are given by Eqns. (2.14) and (2.19) (with skew invariant $w = 0$ for this 2D geometry) and with $G(\theta)$ defined by Eqn. (3):

$$R(t) = \{n_o^2 t^2 + 2R_1 \dot{R}_1 t + R_1^2\}^{1/2} \quad (8)$$

$$R^4 \dot{\theta}^2 = \kappa - n_o^2 N_o^2 G(\theta) \quad (9)$$

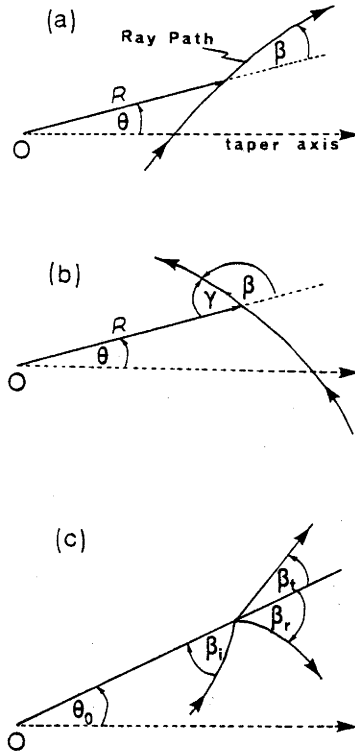


Figure 2 Ray angle, β , at a point (R, θ) on the ray path. In (a) the ray propagates away from the apex, and in (b) towards the apex. In the latter case it is more convenient to refer to the complementary ray angle, $\gamma = \pi - \beta$, as shown. In (c) the ray angles are shown as a ray is reflected and refracted at the taper interface.

Recall that κ is an invariant for each ray and is defined by

$$\kappa \equiv R^2 (n_0^2 - \dot{R}^2) \quad (10)$$

It is important to note that κ remains invariant regardless of reflection or refraction at the taper interface ($\theta = \pm\theta_0$). In fact, using Eqn. (6) to write

$$\kappa = R^2 (n_0^2 - n^2 \cos^2 \beta) \quad (11)$$

we see that the requirement that the reflected and refracted rays conserve κ is equivalent to the law of reflection and Snell's law of refraction, respectively (see Fig. 2(c)).

The behavior of rays inside the taper follows from Eqns. (8) and (9), and is described in the next section.

4.4 RAY PROPAGATION

4.4.1 Ray Path: R Variable

The behavior of the R variable is described by Eqn. (8): as t ranges from $-\infty$ to $+\infty$, R monotonically decreases from infinity, reaches a minimum, and then monotonically increases to infinity. This is depicted schematically in Fig. 3 where we make the distinction between two cases: $\kappa > 0$ (Fig. 3(a)) and $\kappa \leq 0$ (Fig. 3(b)).

When $\kappa > 0$ (see Fig. 3(a)) the radial behavior is the same as for the tapered quadratic medium of chapter 2. In particular, a circular ray caustic:

$$R_{\min} = \sqrt{\kappa} / n_0 \quad (12)$$

determines the minimum value of R (see Eqn. (2.22)). Thus a ray propagating inside the taper is turned back on itself, whereas a ray in the uniform cladding must graze the circle $R = R_{\min}$.

From Fig. 3(b) we see that rays with $\kappa \leq 0$ do not have a radial caustic: such rays must pass through the origin. Note that the refractive index distribution (Eqn. (1)) has a singularity at the origin (it is infinite at $R=0$) and is therefore unphysical at this point. This is responsible for the meaningless interpretation of R beyond this point (R becomes imaginary as shown in Fig. 3(b)). This problem can obviously be avoided if we restrict attention to the ray paths in any section of the medium which excludes the origin. Nevertheless, comparison with the modal

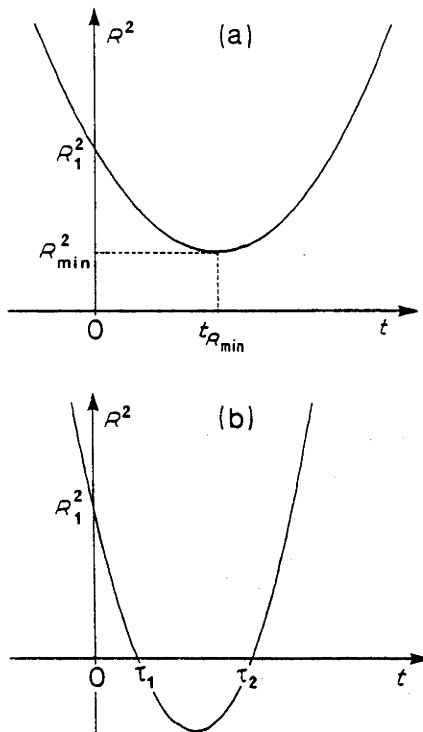


Figure 3 Schematic plot of $R^2(t)$ for the two cases (a) $K > 0$, and (b) $K < 0$. In (a) the ray reaches its closest approach to the origin, $R_{min} = K^{1/2}/n_0$, at $t = -R_1 R_1/n_0^2$. In (b) the ray passes through the origin at $t = (-R_1 R_1 \pm (-K)^{1/2})/n_0^2$ (labelled τ_1 and τ_2 in the figure).

fields given by Marcatili² will later show that this point is of significance when interpreting the radial dependence of the fields.

From Eqn. (11) we obtain the range of values associated with the ray invariant:

$$\begin{aligned} -n_0^2 N_0^2 &\leq \kappa < \infty && \text{for rays inside the taper} \\ 0 &\leq \kappa < \infty && \text{for rays outside the taper} \end{aligned} \quad (13)$$

Two classes of rays - bound and refracting - may propagate inside the taper. It is evident from Eqn. (13), and the invariance of κ , that rays are classified as follows:

$$\begin{aligned} -n_0^2 N_0^2 &\leq \kappa < 0 && : \text{ Bound Rays} \\ 0 &< \kappa < \infty && : \text{ Refracting Rays} \end{aligned} \quad (14)$$

This conclusion could be reached more directly from Snell's law, which gives the condition for total internal reflection as $|\cos\beta| > 1 / \sqrt{(1 + N_o^2/R^2)}$ where β and R refer to the position where the ray touches the interface. Using Eqn. (6), this condition simplifies to $\kappa < 0$.

Thus we have classified the rays according to the value of the invariant, κ . Refracting rays have a circular caustic of radius $R_{\min}$, and so a ray inside the taper coming in from infinity will turn back on itself when it touches the caustic, and then recede from the apex. Whenever the ray touches the interface it bifurcates, and the refracted component must graze the circular caustic (or its back-projection must if the ray is propagating in the opposite direction). Thus the refracting ray loses power very rapidly. Bound rays are trapped inside the taper and propagate through to the apex. The complete picture of ray propagation in the taper follows from an analysis of the θ variable, given in the next section.

4.4.2 Ray Path : θ Variable

The angular behavior is found from Eqn. (9). Inside the taper $G(\theta) = -1$ (Eqn. (3)) and hence:

$$R^4 \dot{\theta}^2 = \kappa + n_o^2 N_o^2 \quad (15)$$

From Eqn. (4) and elementary geometry $\dot{\theta} = n \, d\theta/ds = (n/R)\sin\beta$ and hence

$$nR \sin|\beta| = \sqrt{(\kappa + n_o^2 N_o^2)} = \text{constant} \quad (16)$$

Equation (16) expresses the invariance of κ inside the taper in an alternative form. (Sometimes called Bougers formula³, $nR \sin|\beta|$ is always conserved in a spherically symmetric refractive index, $n = n(R)$.) However, unlike κ which is an invariant in a global sense, $nR \sin|\beta|$ is an invariant in a somewhat more restricted sense. This follows from the fact that although a reflected ray maintains the same value of $nR \sin|\beta|$, a refracted ray does not.

Consider a ray propagating towards the apex of the taper. We use the complementary ray angle, γ (see Fig. 2(b)), and from Eqn. (16):

$$|\gamma| = \sin^{-1} \{ (\kappa + n_o^2 N_o^2) / (n_o^2 R^2 + n_o^2 N_o^2) \}^{1/2} \quad (17)$$

Figure 4 shows $|\gamma|$ as a function of R , and reveals that $|\gamma|$ steadily increases as the ray approaches the taper apex. Thus, the ray strikes the interface at a progressively steeper angle (reflection merely reverses the sign of γ). (The exception is the special case $\kappa = -n_o^2 N_o^2$ for which $\gamma \equiv 0$, corresponding to straight, radial trajectories.) Note that due to the $1/R^2$ refractive index grading, the critical angle for total internal reflection, $|\gamma_c|$, increases in precisely the same way:

$$|\gamma_c| = \sin^{-1} \{ n_o N_o / \sqrt{n_o^2 R^2 + n_o^2 N_o^2} \} \quad (18)$$

Comparing Eqns (17) and (18) shows that $|\gamma| < |\gamma_c|$ for bound rays. This explains how a ray can remain trapped all the way down to the apex.

Figure 4 also gives the behavior of rays far from the apex. Note that $\gamma \rightarrow 0$ as $R \rightarrow \infty$ and thus the ray trajectories tend asymptotically to straight lines. This was pointed out in chapter 2, and is a consequence of the fact that the grading is weak for large R .

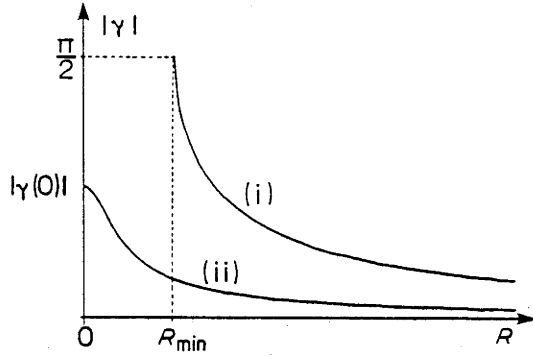


Figure 4 Modulus of the complementary ray angle as a function of the R -variable for the trajectory inside the taper. Curve (i) refers to refracting rays, while curve (ii) refers to bound rays. At the instant the bound rays reach the origin the ray angle has the value $|\gamma(0)| = \arcsin [(K + n_0^2 N_0^2)/n_0^2 N_0^2]^{1/2}$.

As the ray approaches the origin it begins to oscillate, reflecting off the taper interfaces. Substituting $\dot{\theta} = \dot{R} d\theta/dR$ in Eqn. (15) we obtain

$$\frac{d\theta}{dR} = \pm \frac{\sqrt{(\kappa + n_0^2 N_0^2)}}{R\sqrt{(n_0^2 R^2 - \kappa)}} \quad (19)$$

and the rate of oscillation increases as the ray approaches the apex. A refracting ray ($\kappa > 0$) will touch the caustic at $R_{\min}$ (Eqn. (12)) turn around and leave the taper (thus the singularity in $d\theta/dR$ at $R = R_{\min}$). On the other hand, a bound ray ($\kappa < 0$) approaching the origin has its rate of oscillation increase without limit.

Integrating Eqn. (19) gives θ explicitly:

$$\theta(R) = \pm \begin{cases} -[(N_0/\sigma)^2 - 1]^{1/2} \operatorname{cosech}^{-1}(R/\sigma) + C_1, & -n_0^2 N_0^2 < \kappa < 0 \\ [(N_0/R_{\min})^2 + 1]^{1/2} \sec^{-1}(R/R_{\min}) + C_2, & \kappa > 0 \end{cases}$$

(20)

where

$$\sigma \equiv (1/n_o) \sqrt{|\kappa|} \quad (21)$$

and the $\pm$ signs refer to alternate ray path segments for which $d\theta/dR \gtrless 0$ (i.e. between reflections). The constants C_1 and C_2 are determined from initial conditions.

The characteristic behavior of the bound and refracting rays is summarized in Fig. 5.

4.4.3 Changing the External Refractive Index

We now briefly discuss the changes that occur when the uniform cladding index is decreased or increased from its original value n_o . To this end, consider the index distribution

$$n^2 = \begin{cases} n_o^2 \{1 + N_o^2 / R^2\} , & |\theta| < \theta_o \\ n_1^2 , & \theta_o < |\theta| < \pi \end{cases} \quad (22)$$

where $n_1 \neq n_o$. The ray path equation is no longer separable throughout the entire region, and κ (Eqns. (10)-(11)) is no longer a global invariant. However, for the part of the trajectory lying inside the taper, κ remains an invariant, and is preserved following reflection at the interface. Thus we still use κ to identify rays inside the taper.

Snell's law shows that a ray striking the interface a distance R from the apex will be totally internally reflected if $\kappa < (n_o^2 - n_1^2)R^2$ and refracted otherwise. This leads to the characteristic radius

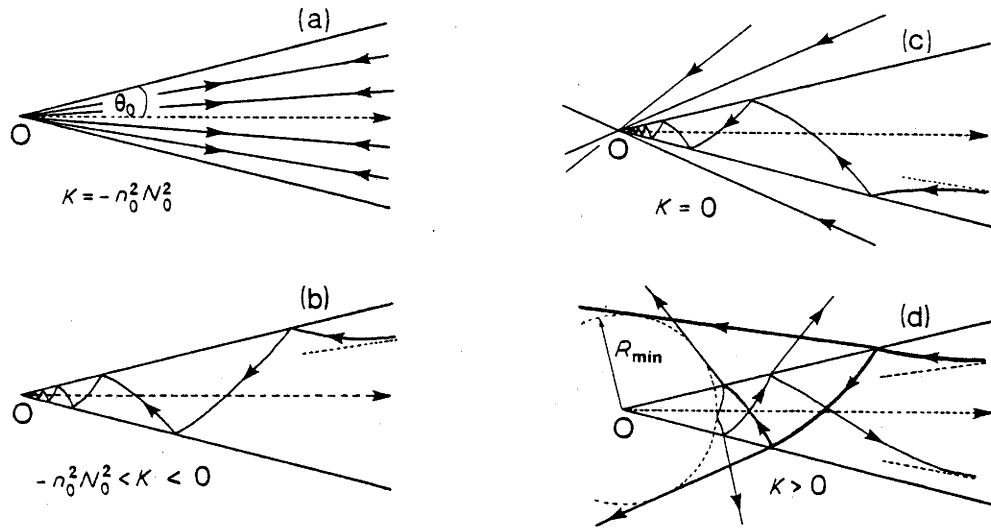


Figure 5 Ray paths in the tapered waveguide, classified according to the ray invariant K . Bound rays are shown in (a) to (c), refracting rays in (d). Note that the refracting rays lose power rapidly and are almost negligible after only a couple of reflections [as indicated schematically by the line thickness in (d)].

$$R_c = \{\kappa / (n_0^2 - n_1^2)\}^{1/2} \quad (23)$$

We consider the two cases $n_1 < n_0$ and $n_1 > n_0$ separately -

When $n_1 < n_0$, rays with $-n_0^2 N_0^2 < \kappa < 0$ always totally internally reflect and thus remain bound. On the other hand, a ray with $\kappa > 0$ will totally internally reflect when $R > R_c$ but will refract when $R < R_c$. Thus we can consider R_c to be a cut-off radius marking the appearance of radiation. Note that taking the limit $N_0 \rightarrow 0$ gives a step-index taper, and the bound rays no longer exist.

When $n_1 > n_0$, rays with $\kappa > 0$ always refract on striking the interface. However, a ray with $-n_0^2 N_0^2 < \kappa < 0$ will totally internally reflect when $R < R_c$ but will refract when $R > R_c$. The latter is a consequence of the fact that at large R the refractive index inside the taper becomes less than in the cladding - giving an antiwaveguiding structure.

4.5 RAY COMPRESSION

The bound rays are trapped inside the taper and become concentrated in a progressively smaller region as they propagate towards the apex. The physical principles of ray compression have been investigated for several years⁴, with widespread applications in radiation detectors⁵ and solar collectors⁶, so it is interesting to study the strictly adiabatic taper in this light.

Figure 6 gives a schematic illustration of the concentrator problem. The natural geometry of the taper suggests we choose for the entrance aperture a radial arc of radius R_1 and length $(2\theta_o) R_1 \equiv L_1$. Rays are launched from every point of the entrance aperture into the angular range $|\gamma| < \gamma_i$ as shown in the figure. We call γ_i the acceptance angle of the concentrator. Concentrator theory aims to direct all the given rays on the entrance aperture over to an exit aperture at $R = R_2$ of length $(2\theta_o) R_2 \equiv L_2 < L_1$. The concentration ratio is then defined as $C \equiv L_1 / L_2$.

The concentrator we are describing can be viewed in two equivalent ways. The ray components $R(t)$, $\theta(t)$ are given by Eqns. (8) and (20). Introducing a z -axis normal to the R - θ plane (i.e. R - θ - z are cylindrical polars) and assuming the refractive index is independent of z , gives a concentrator with trough geometry, and $z(t) = \dot{z}(0)t + z(0)$ completes the ray path description. Alternatively, introducing an azimuthal angle ϕ (i.e. R - θ - ϕ are spherical polars) and assuming the index is independent of ϕ , gives a concentrator with conical geometry, and $R(t)$, $\theta(t)$, $\phi \equiv \text{const.}$ describes the meridional rays. The entrance and exit apertures are cylindrical in the former case, and spherical in the latter.

From Eqns. (14) and (17) it follows that all rays launched from the entrance aperture with complementary ray angle in the range

$$0 < |\gamma| < \sin^{-1}\{N_o / \sqrt{(N_o^2 + R_1^2)}\} \quad (24)$$

are bound. Thus setting the acceptance angle as:

$$\gamma_i = \sin^{-1}\{N_o / \sqrt{(N_o^2 + R_1^2)}\} \quad (25)$$

we see that all the rays will be transported, without loss, over to the exit aperture, and any arbitrarily large concentration ratio may be obtained simply by making R_2 small enough. (Of course we must keep in mind that the geometrical optics approximation will break down when the refractive index gradient becomes too large, near the taper apex - see section 1.5.) At the exit aperture the rays will have an angular spread defined by $|\gamma| < \gamma_f$ where

$$\gamma_f = \sin^{-1}\{N_o / \sqrt{(N_o^2 + R_2^2)}\} > \gamma_i \quad (26)$$

This is illustrated schematically in Fig. 6. In the sense that all the given rays are directed to the exit aperture without loss, the taper is a perfect concentrator.

The above is consistent with the fundamental physical limitation on ray compression imposed by Liouville's theorem of phase space conservation. This limitation may be stated simply as⁷:

$$n_1 L_1 \sin \gamma_i \leq n_2 L_2 \quad (27)$$

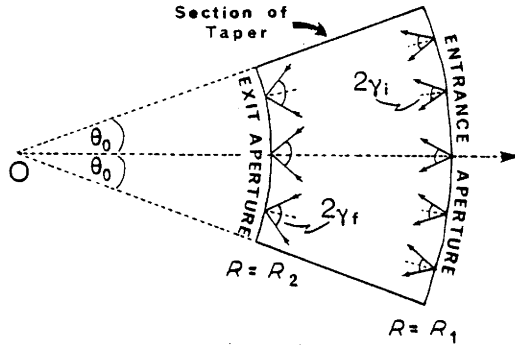


Figure 6 The concentrator problem. Rays are launched into the taper from every point of the entrance aperture, within a half-angle γ_i . They arrive at the exit aperture within a half-angle γ_f .

for a 2D concentrator of arbitrary optical design (see Fig. 7). Here L_1 and L_2 are the lengths of the entrance and exit apertures respectively, and n_1 and n_2 are the refractive indices there. (The refractive index is assumed constant over both apertures.) Equation (27) shows that given an acceptance angle, γ_i , and refractive indices n_1 and n_2 , then there is a theoretical maximum concentration ratio

$$C_{\max} = n_2 / n_1 \sin \gamma_i \quad (28)$$

that can never be exceeded by any concentrator built from refracting or reflecting elements.

Clearly $C_{\max}$ can be made arbitrarily large simply by increasing the refractive index at the exit aperture, n_2 . This situation is occurring with the strictly adiabatic taper as R_2 is decreased, due to the $1/R^2$ grading within the taper. The question of interest is whether or not the theoretical maximum concentration is actually reached. Using Eqns. (1), (25) and (28) we find that

$$C = C_{\max} / \sqrt{(1 + R_2^2 / N_o^2)} \quad (29)$$

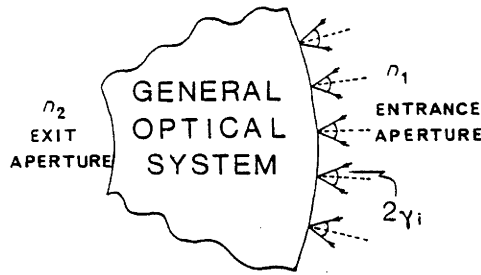


Figure 7 An arbitrary two-dimensional concentrator with plane-polar geometry. The entrance and exit apertures are assumed to be circular arcs of length L_1 and L_2 , respectively. The refractive index is assumed to be constant over each aperture, with values n_1 and n_2 .

where $C = L_1/L_2$ is the actual concentration ratio of the taper. Thus the theoretical concentration maximum is only reached in the limit $R_2 \rightarrow 0$. The taper (which concentrates the rays by a combination of reflection and refraction) is less effective as a concentrator than designs which do achieve the theoretical maximum: e.g. the Luneburg lens based on refracting optics, and the trough compound parabolic concentrator based on reflecting optics⁴.

In some applications the concentrator should ideally reject all rays launched outside of the acceptance angle, γ_i ⁴. This is clearly not the case with the strictly adiabatic taper, since a minute amount of power will be collected from the refracting rays. From Eqns. (12) and (17) it follows that rays launched into the taper with

$$\gamma_i < |\gamma| < \sin^{-1} \{ (N_o^2 + R_2^2) / (N_o^2 + R_1^2) \}^{1/2} \quad (30)$$

will reach the exit aperture after suffering lossy reflections at the taper interface. Any rays launched with $\sin^{-1} \{ (N_o^2 + R_2^2) / (N_o^2 + R_1^2) \}^{1/2} < |\gamma| < \pi/2$ will encounter the caustic at $R_{\min}$ and be turned back, before reaching the exit aperture. Thus, a small amount of power launched outside the acceptance angle but within the angular width

$$\Delta|\gamma| = \sin^{-1}\{(N_o^2 + R_2^2) / (N_o^2 + R_1^2)\}^{1/2} - \sin^{-1}\{N_o^2 / (N_o^2 + R_1^2)\}^{1/2} \quad (31)$$

will be received at the exit aperture. Note that $\Delta|\gamma| \ll \gamma_i$ when $R_2 \ll N_o$.

Strictly adiabatic tapers may find application as light concentrators in micro-optic⁸ systems (see section 1.4.1 of the introductory chapter). But either way, the simple example given here is interesting from a theoretical viewpoint because it can be solved exactly, and provides a further example of concentrator design.

4.6 COMPARISON WITH MODAL FIELDS

It is interesting to compare the properties of the ray solution with the modal fields derived by Marcatili². Assuming an implicit time dependence $\exp(-i\omega t)$ where $\omega = ck$ is the monochromatic source frequency, the Helmholtz equation can be solved exactly by separation of variables (see chapter 3). (The electromagnetic field can be constructed from the solutions of the Helmholtz equation under the well-known weak guidance assumption⁹, as discussed in section 1.5.)

Writing the modal field in the separated form:

$$\psi(R, \theta) = A(R)B(\theta) \quad (32)$$

it is found that the radial and angular dependence of the even modes (the odd modes have similar properties) can be expressed as:

$$\text{BOUND MODES:} \quad -k^2 n_o^2 N_o^2 < \nu < 0 \quad (33)$$

$$A(R) = \begin{cases} J_{i\sqrt{|\nu|}}(kn_o R) \\ Y_{i\sqrt{|\nu|}}(kn_o R) \end{cases} \quad (34)$$

$$B(\theta) = \begin{cases} \cos\{\theta\sqrt{\nu + k^2 n_o^2 N_o^2}\} & , \quad |\theta| < \theta_o \\ \cos\{\theta_o\sqrt{\nu + k^2 n_o^2 N_o^2}\} \frac{\cosh\sqrt{|\nu|}(\pi - |\theta|)}{\cosh\sqrt{|\nu|}(\pi - \theta_o)} & , \quad \theta_o < |\theta| < \pi \end{cases} \quad (35)$$

where ν satisfies the eigenvalue equation:

$$\sqrt{\nu + k^2 n_o^2 N_o^2} \tan\{\theta_o\sqrt{\nu + k^2 n_o^2 N_o^2}\} = \sqrt{|\nu|} \tanh\sqrt{|\nu|}(\pi - \theta_o) \quad (36)$$

$$\text{RADIATION MODES:} \quad \nu > 0 \quad (37)$$

$$A(R) = \begin{cases} J_{\sqrt{\nu}}(kn_o R) \\ Y_{\sqrt{\nu}}(kn_o R) \end{cases} \quad (38)$$

$$B(\theta) = \begin{cases} \cos\{\theta\sqrt{\nu + k^2 n_o^2 N_o^2}\} & , \quad |\theta| < \theta_o \\ \cos\{\theta_o\sqrt{\nu + k^2 n_o^2 N_o^2}\} \frac{\cos\sqrt{\nu}(\pi - |\theta|)}{\cos\sqrt{\nu}(\pi - \theta_o)} & , \quad \theta_o < |\theta| < \pi \end{cases} \quad (39)$$

where ν satisfies the eigenvalue equation:

$$\sqrt{(\nu + k^2 n_o^2 N_o^2)} \tan\{\theta_o \sqrt{(\nu + k^2 n_o^2 N_o^2)}\} = -\sqrt{\nu} \tan\sqrt{\nu} (\pi - \theta_o) \quad (40)$$

Note that the radial dependence is described by some combination of Bessel functions¹⁰, the exact form of which depends upon the boundary conditions (we refer the reader back to section 3.5).

Links between the ray and modal descriptions will now be discussed. In section 4.6.1 we compare radiation modes and refracting rays, and in section 4.6.2 guided modes and bound rays.

4.6.1 Radiation Modes and Refracting Rays

Consider the situation with all sources located in the region $R > R_o$ of the taper, and we wish to determine the field in the region $R < R_o$ (see Fig. 8a). We assume that the taper extends all the way down to the origin. In this case we must choose for the radial dependence of the radiation modes (Eqn. (38)):

$$A(R) = \frac{J_{\sqrt{\nu}}(kn_o R)}{\sqrt{\nu}} \quad (41)$$

so that the modal fields are bounded everywhere in the region $R < R_o$, and are vanishingly small at the apex of the taper. With this radial dependence, the modes describe standing wave fields, in agreement with the geometrical optics picture of rays propagating into the taper and reflecting back out after encountering the circular ray caustic of radius $R_{\min}$. Recall from section 3.5.1, that the wave caustic is approximately located where the argument equals the order of the Bessel function in Eqn. (41). If $\tilde{R}_\nu$ denotes the radius of the caustic, then

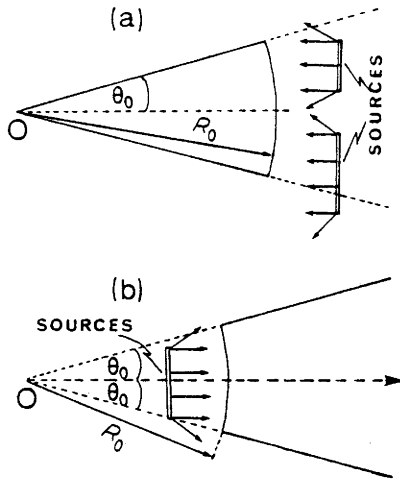


Figure 8 Propagation in the complete taper. In (a) propagation towards the apex is illustrated, and in (b) that away from apex. The details of the medium in the regions $R > R_0$ in (a), and $R < R_0$ in (b) are unimportant.

$$\tilde{R}_v = \sqrt{v} / kn_0 \quad (42)$$

A family of refracting rays, all having the same value invariant κ , may be associated with the mode of eigenvalue v by matching the ray and wave caustics. Setting $R_{\min} = \tilde{R}_v$ gives:

$$\kappa = v / k^2 \quad (43)$$

The above relation gives the criterion for a ray family to correspond to a mode, in analogy with Eqn. (3.34) of chapter 3.

For completeness, we comment on the situation depicted in Fig. 8(b): Assuming all sources in the region $R < R_0$, we are interested in the field in the region $R > R_0$, with the taper extending out to infinity. In this case the radial dependence of the modal fields is given by a Hankel function¹⁰:

$$A(R) = \frac{H^{(1)}_v(kn_0 R)}{\sqrt{v}} \quad (44)$$

This follows from the asymptotic form of Hankel functions for large R :

$$\mathcal{H}_{\sqrt{v}}^{(1),(2)}(k n_o R) \sim (2/\pi k n_o R)^{1/2} \exp\{\pm i[k n_o R - \sqrt{v}(\pi/2) - \pi/4]\} \quad (45)$$

(where $\pm$ refers to $\mathcal{H}_{\sqrt{v}}^{(1)}$ and $\mathcal{H}_{\sqrt{v}}^{(2)}$ respectively) and the obvious physical requirement that there be no backward propagating field at infinity.

4.6.2 GUIDED MODES AND BOUND RAYS

The radial dependence of the bound modes is given by a linear combination of Bessel functions of real argument and imaginary order (see Eqn.(34)). It is interesting to note that Bessel functions of this type occur very infrequently in mathematical physics (see, for example, Bocher¹¹ for some discussion). The asymptotic behaviour of these Bessel functions, as the argument tends to zero, is given by¹⁰:

$$\Gamma(1 + i\mu) J_{i\mu}(x) \sim \exp[i\mu \log(x/2)] \quad (46)$$

and a similar expression for $Y_{i\mu}(x)$. Both Bessel functions are oscillatory near the origin, with the rate of oscillation increasing without limit as the origin is approached. The corresponding behaviour of bound rays near the apex of the taper was discussed in section 4.4.2. There it was noted that a bound ray oscillates as it propagates towards the origin with its rate of oscillation increasing without limit as the apex is approached.

The bound modes have a singularity at the origin because they do not vanish at this point, and the refractive index is unphysical there. This corresponds to bound rays reaching the apex of the taper. Thus it is

physically meaningless to use this model to describe propagation of bound modes towards the apex of the complete taper (Fig.8(a)). The bound modes can only be used in a section of the taper which excludes the origin. For example, using the bound modes to describe propagation up the complete taper (away from the apex) as in Fig. 8(b), we find that the appropriate radial dependence is:

$$A(R) = \frac{H^{(1)}_{i\sqrt{|v|}}(kn_0 R)}{i\sqrt{|v|}} \quad (47)$$

This follows because Eqn. (45) giving the asymptotic form of the Hankel functions for large R is still valid when $\sqrt{v}$ is replaced by $i\sqrt{|v|}$, so the argument used to derive Eqn. (44) applies here as well.

4.7 3D TAPER

We now derive the exact ray solution for the 3D taper formed by rotating the planar taper about its axis of symmetry.

The refractive index distribution is given by Eqns. (1)-(3), but with $R-\theta-\phi$ as spherical polar coordinates, ϕ being the azimuthal angle (see Fig. 9). The taper is now a solid cone of graded-index material surrounded by a uniform medium of refractive index n_0 . The refractive index inside the cone depends upon the radial variable only, in accordance with Eqn. (1).

The ray trajectories of the planar taper are the meridional ray paths for the 3D taper. Consequently, we concern ourselves with the characteristics of the skew rays, and show that this results in the introduction of tunnelling rays into the class of leaky rays.

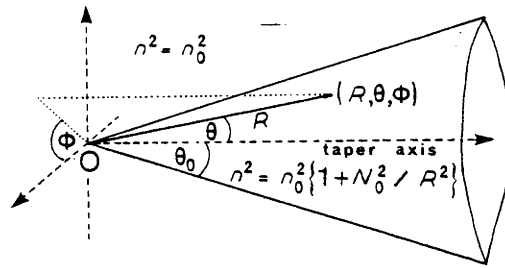


Figure 9 Co-ordinate system for three-dimensional taper. Here R - θ - ϕ are spherical polar co-ordinates, with ϕ the azimuthal angle. We follow the usual convention: $R \geq 0$, $0 \leq \theta \leq \pi$, $0 \leq \phi \leq 2\pi$.

4.7.1 Ray Equations

It is shown in chapter 2 that the radial component of the trajectory is given by Eqn. (8) and that κ (Eqs. (10)-(11)) is a ray invariant. This is true for the skew ray paths as well as for the meridional trajectories. The angular components of the ray path are found from (see Eqns. (2.14) and (2.15)):

$$R^4 \dot{\theta}^2 = \kappa - n_0^2 N_0^2 G(\theta) - w^2 / \sin^2 \theta \quad (48)$$

$$w = R^2 \dot{\phi} \sin^2 \theta \quad (49)$$

where $G(\theta)$ is defined in Eqn. (3). Here w is the skew invariant. Meridional rays have $w = 0$ while skew rays correspond to $w \neq 0$. Equation (49) shows that ϕ is simply a monotonically increasing or decreasing function of the parameter t , describing the rotation of the ray about the axis of the taper. The sign of w gives the sense of the rotation.

The fact that the radial component of the trajectory is given by Eqn. (8) for both meridional and skew rays allows us to refer to the discussion in section 4.4.1 with only minor changes. In particular, rays with $\kappa > 0$ have a caustic which is now a spherical surface of radius $R_{\min}$ (Eqn. (12)) centred on the apex of the conical taper. Rays with $\kappa < 0$ propagate through to the apex.

Rays in the 3D taper are labelled by the two invariants, w and κ , and classification as bound or leaky follows from Eqn. (48).

4.7.2 The θ Equation

The θ variable is obtained by solving Eqn. (48). Noting that the RHS of Eqn. (48) must be non-negative, and requiring that the taper angle $\theta_0 < \pi/2$, we are led to the three possible cases illustrated in Fig. 10(a)-(c).

From Fig. 10 it is clear that we must have

$$\kappa + n_0^2 N_0^2 > w^2 / \sin^2 \theta_0 \quad (50)$$

for a ray to exist inside the taper. Hereon we shall assume that this is in fact the case. A ray inside the taper has θ bounded as follows:

$$\theta_{\min} < \theta < \theta_0 \quad (51)$$

where θ_0 corresponds to the taper interface, and $\theta_{\min}$ is an inner ray caustic defined by

$$\theta_{\min} = \sin^{-1} \{ |w| / \sqrt{(\kappa + n_0^2 N_0^2)} \} \quad (52)$$

The corresponding range that θ can assume in the region outside the taper is different for three distinct cases to be considered in turn:

CASE I: $\kappa > w^2 / \sin^2 \theta_0$

In this case the ray outside the taper has (see Fig. 10(a)):

$$\theta_0 < \theta < \theta_{\text{MAX}}^{\text{RAD}} \quad (53)$$

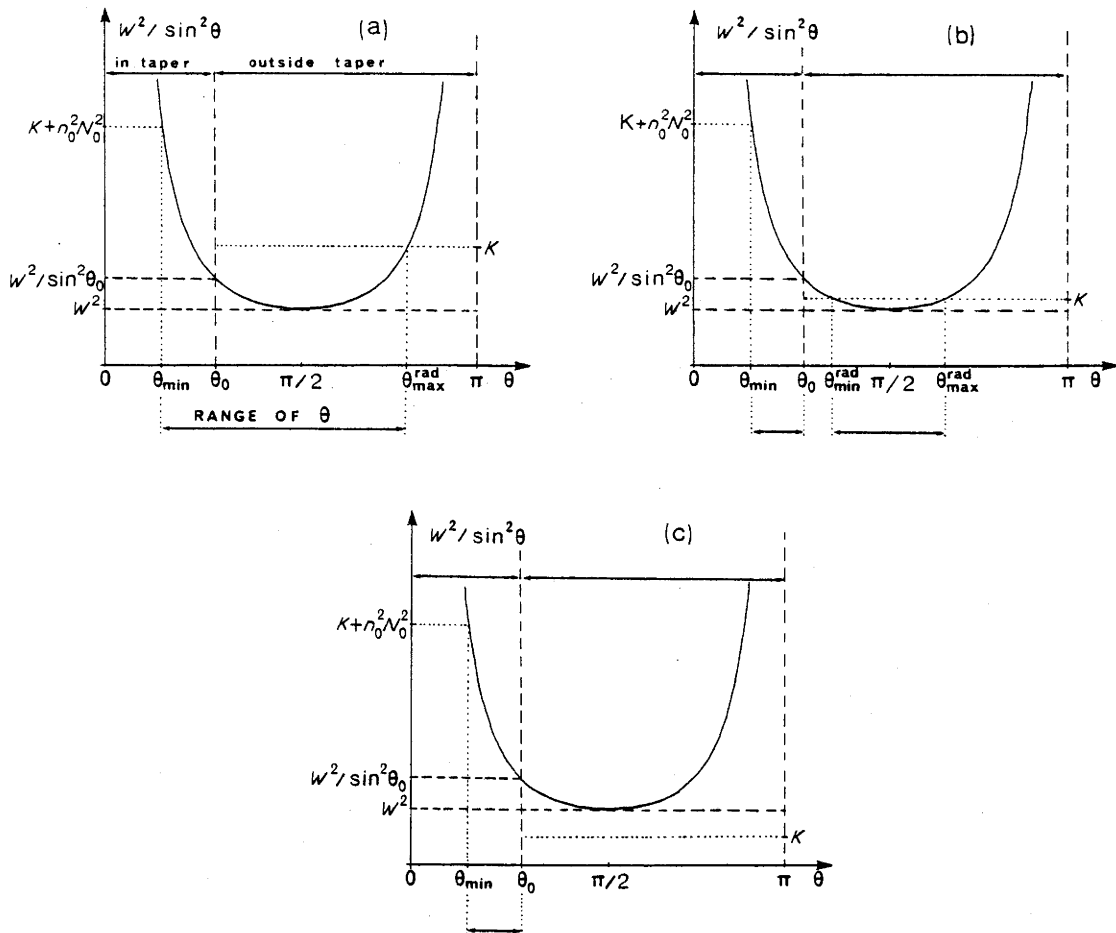


Figure 10 Range of θ satisfying the requirement that the right-hand side of Equation 48 be non-negative, for the three cases: (a) $K \geq W^2/\sin^2\theta_0$, (b) $W^2 \leq K \leq W^2/\sin^2\theta_0$ and (c) $-n_0^2 N_0^2 \leq K < W^2$. Thus, we have refracting rays, tunnelling rays and bound rays in (a), (b) and (c), respectively.

where $\theta_{\text{MAX}}^{\text{RAD}}$ is an outer radiation caustic. This situation corresponds to refracting rays, since the ray outside the taper originates on the taper interface.

The outer radiation caustic is given explicitly by

$$\theta_{\text{MAX}}^{\text{RAD}} = \pi - \sin^{-1} \left\{ |w| / \sqrt{\kappa} \right\} \quad (54)$$

and from the geometry we see that it corresponds to a cone lying behind the taper.

CASE II: $w^2 < \kappa < w^2/\sin^2\theta_0$

In this case the ray outside the taper has (see Fig. 10(b)):

$$\theta_{\text{MIN}}^{\text{RAD}} < \theta < \theta_{\text{MAX}}^{\text{RAD}} \quad (55)$$

where $\theta_{\text{MAX}}^{\text{RAD}}$ is given by Eqn. (54) and $\theta_{\text{MIN}}^{\text{RAD}}$ is an inner radiation caustic given by

$$\theta_{\text{MIN}}^{\text{RAD}} = \sin^{-1} \{ |w| / \sqrt{\kappa} \} \quad (56)$$

From the spherical polar geometry and a comparison of Eqns (54) and (56) we see that the inner radiation caustic may be viewed as an extension of the outer radiation caustic into the adjoining half-space.

Note that rays can propagate in two distinct regions, separated by a forbidden zone. In wave theory, these two regions will be linked by an evanescent field, and power transfer between the two regions will proceed by a tunnelling effect - a process analogous to frustrated total internal reflection, and quantum mechanical tunnelling. Thus we have the class of rays known as tunnelling rays. Such rays leak extremely slowly compared to refracting rays⁹.

CASE III: $\kappa < w^2$

In this case (Fig. 10(c)) no ray can exist outside the taper and clearly any ray satisfying this condition will be a bound ray.

4.7.3 Ray Path Characteristics

All the rays inside the taper have been classified according to the values of the ray path invariants, w and κ . This is summarized in Fig. 11.

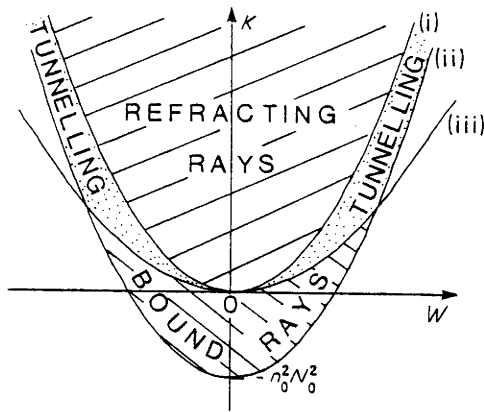


Figure 11 Classification of rays according to the two ray invariants, W and K . The curves shown are (i) $K = W^2 / \sin^2 \theta_0$, (ii) $K + n_0^2 N_0^2 = W^2 / \sin^2 \theta_0$ and (iii) $K = W^2$. Note that meridional rays correspond to the K -axis (that is, the line $W = 0$).

Note from Fig. 11 that some bound rays have $\kappa > 0$, and others have $\kappa < 0$. Thus, in contrast to the meridional case, some bound skew rays have a caustic of radius $R_{\min}$ and cannot propagate through to the taper apex.

The θ variable of the ray inside the taper oscillates within the limits set by Eqn. (51). We can show from Eqn. (48) that

$$\frac{d\theta}{dR} = \pm \frac{\sqrt{(\kappa + n_0^2 N_0^2 - w^2 / \sin^2 \theta)}}{R \sqrt{(n_0^2 R^2 - \kappa)}} \quad (57)$$

in analogy with Eqn. (19). Here the $\pm$ signs refer to alternate segments of the ray path between reflections off the taper interface (at $\theta = \theta_0$) and inner caustic (at $\theta = \theta_{\min}$). Once again, the rate of oscillation increases as the ray propagates toward the apex of the taper.

Finally, Eqn. (57) can be integrated to obtain θ explicitly:

$$\cos \theta = \begin{cases} \pm \xi \sin \{ [(N_0 / \sigma)^2 - 1]^{1/2} \operatorname{cosech}^{-1} (R / \sigma) + C_1 \}, & -n_0^2 N_0^2 \leq \kappa < 0 \\ \pm \xi \sin \{ [(N_0 / R_{\min})^2 + 1]^{1/2} \sec^{-1} (R / R_{\min}) + C_2 \}, & \kappa > 0 \end{cases} \quad (58)$$

where

$$\xi = \{(\kappa + n_o^2 N_o^2 - w^2) / (\kappa + n_o^2 N_o^2)\}^{1/2} \quad (59)$$

and σ is defined in Eqn. (21).

4.7.4 The ϕ Equation

For completeness we also present the exact solution for the ϕ component (see Eqn. (49)). It is most convenient to express ϕ as a function of θ , by using Eqn. (48) with Eqn. (49). This leads to:

$$\phi = \pm \chi \int \frac{d\theta}{\sin\theta \sqrt{(\sin^2\theta - \chi^2)}} \quad (60)$$

where $\chi = w / \sqrt{(\kappa + n_o^2 N_o^2)}$.

4.8 CONCLUSION

Ray optics gives a more physical description of the guidance of light in strictly adiabatic tapers. For bound rays launched down the taper, the $1/R^2$ grading raises the critical angle for total internal reflection in just such a way as to compensate for the ever steepening ray angle at each reflection. The conical geometry leads to compression of the bound rays, and to the alternative description of the taper as a non-imaging concentrator. Although perfect in the sense that all rays (launched within the acceptance angle) are transported from entrance to exit aperture without loss, it does not attain the theoretical maximum concentration established by Liouville's theorem.

Later (chapter 7) we return to this example and investigate the application of this taper as a low-loss connector of waveguides with unequal cross-section. The main source of loss will be discussed, and quantified for the case of fundamental mode transmission.

4.9 REFERENCES

1. D. Marcuse, "Light Transmission Optics," (Van Nostrand Reinhold, New York, 1982).
2. E.A.J. Marcatili, "Dielectric Tapers with Curved Axes and No Loss," IEEE J. Quantum Electron. QE-21, 307 (1985).
3. S. Cornbleet, "Microwave and Optical Ray Geometry", (Wiley, New York, 1984).
4. W.T. Welford and R. Winston, "The Optics of Nonimaging Concentrators", (Academic, New York, 1978).
5. H. Hinterberger and R. Winston, "Efficient Light Coupler for Threshold Cerenkov Counters," Rev. Sci. Instrum. 37, 1094 (1966).
6. R. Winston and H. Hinterberger, "Principles of Cylindrical Concentrators for Solar Energy," Solar Energy 17, 255 (1975).
7. R. Winston, "Light Collection within the Framework of Geometrical Optics," J. Opt. Soc. Am. 60, 245 (1970).

8. K. Iga, Y. Kokubun and M. Oikawa, "Fundamentals of Microoptics", (Academic, New York, 1984).
9. A.W. Snyder and J.D. Love, "Optical Waveguide Theory", (Chapman and Hall, London, 1983).
10. M. Abramowitz and I.A. Stegun (eds.), "Handbook of Mathematical Functions", (Dover, New York, 1972).
11. M. Bocher, "On some Applications of Bessel's Functions with Pure Imaginary Index," Ann. Math. 6, 137 (1892).

CHAPTER 5

DIFFRACTION AND SELF-FOCUSING IN GRADED-INDEX MEDIA I:

FORMALISM

5.1 INTRODUCTION

Previous chapters have dealt with the behavior of individual modes (or basis fields) in tapered GRIN media, through the analysis of particular models which are simple enough to reveal the underlying physics. Here, and in the next chapter, we consider diffraction and self-focusing: phenomena involving a large number of modes.

We investigate some effects associated with tapering, by firstly giving a detailed study of scalar diffraction and self-focusing in an untapered quadratic-index medium (this chapter), and then extending the analysis to the case in which the medium is slowly, but otherwise arbitrarily, tapered (chapter 6). In this way we isolate the behavior which develops as a consequence of the tapering.

The content of this chapter is particularly relevant to the problem of image transmission through single GRIN rods¹⁻³ (see also section 1.3). Iga and coworkers^{1,4-6} have obtained an integral expression for the diffracted field in a more general transversely graded medium by using the modal propagation constants found by perturbation analysis, and neglecting terms beyond first order. This is equivalent to the approach used here (see also Sodha and Ghatak⁷) which we establish as an appropriate diffraction formalism which may be applied to any transversely graded, guiding medium. The general formalism is described in section 5.2, and

applied to a quadratically graded medium in section 5.3. This gives the same integral expression obtained by Iga, but we go further, in section 5.4, by giving a detailed study of the structure of the focused field when plane waves are diffracted by an apertured thin lens at the endface of the quadratic medium. (Equivalently, we could consider spherical waves incident on the aperture, without reference to thin lenses.) The field expressions are remarkably simple, and allow us to obtain the critical parameters which characterize the focused field. Explicit formulae are obtained for the intensity half-widths, and the symmetry of the intensity distribution is discussed. We also describe a new focusing effect - the splitting of the intensity maxima. Comparisons are made throughout, with the classic wave-optics problem concerning the structure of the focused field in a uniform medium. In section 5.5 we derive conditions which ensure the validity of the paraxial approximation. Elementary ray optics, in section 5.6, gives a valuable physical interpretation of several results from previous sections. The case of obliquely incident plane waves is analysed in section 5.7 (mathematical details are relegated to an appendix). Finally, the important features of focusing in quadratic-index media are summarized and discussed in section 5.8, and compared with the corresponding results for uniform media. Concluding remarks form section 5.9.

The results are of special interest to the field of visual photoreceptor optics. Pask and Barrell^{8,9} have made a detailed study of the optical factors influencing photoreceptor excitation. Their simplified model involved the focusing of incident light onto the photoreceptors by a thin lens, separated from the photoreceptors by a homogeneous index medium. A more realistic model might involve a weakly graded medium

separating the lens from the photoreceptors, and to a first approximation this grading might be taken to be quadratic. References to this interesting application are made in this chapter.

5.2 GENERAL FORMULATION AND THE PARAXIAL APPROXIMATION

The usual approaches to scalar diffraction calculations; Kirchhoff-Fresnel, Rayleigh-Sommerfeld, Debye etc., apply only to the case of homogeneous media. However, the technique of using a mode expansion and linearizing the propagation constants is quite general and can be applied to any translationally-invariant, guiding medium.

As illustrated in Figure 1, we consider an arbitrary aperture, ξ , in an opaque screen located at $z = 0$. Plane-polar, (r, θ) , coordinates are defined on the aperture plane in the usual way.

Suppose that in the $z > 0$ half-space we have a transversely graded medium:

$$n^2 = n_0^2 (1 - \alpha(r, \theta)) \quad . \quad (1)$$

For convenience, the medium is assumed to be homogeneous (with refractive index $n^2 = n_1^2$) in the $z < 0$ half-space.

We are interested in the diffracted field in the $z > 0$ region, due to some incident field (produced by sources in the $z < 0$ space) striking the aperture. An implicit time-dependence, $\exp[-i\omega t]$, is assumed.

The index profile, Eq. (1), may be viewed as consisting of a grading $(n_0^2 \alpha(r, \theta))$ superimposed on a homogeneous base (n_0^2) . Thus we can speak of the effect, on the diffracted field, of grading the medium, by comparing the diffracted field in the medium defined by Eq. (1) to the analogous field in the homogeneous medium, $n^2 = n_0^2$.

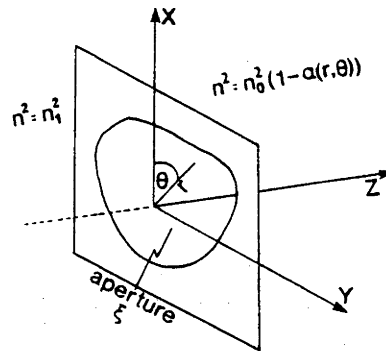


Fig. 1. Coordinate system. The aperture denoted by ξ is located on the $z = 0$ plane of a rectangular coordinate system (x, y, z) . Plane-polar $(r - \theta)$ coordinates are defined on the aperture plane. The medium is graded in the $z > 0$ region and homogeneous in the $z < 0$ region.

The wavelength of light in the homogeneous base, divided by 2π , provides us with a convenient, natural length scale. In this chapter, all lengths will be assumed to be given in units of this length scale. Thus, for example, r and z (along with all other length quantities that will be introduced later) are assumed to have been made dimensionless by scaling in this way.

The exact solution to the Helmholtz equation can be written in the form

$$E(r, \theta, z) = \exp[iz] g(r, \theta, z) \quad , \quad (2)$$

where g satisfies the equation:

$$\partial^2 g / \partial r^2 + (1/r) \partial g / \partial r + (1/r^2) \partial^2 g / \partial \theta^2 - \alpha g + 2i \partial g / \partial z + \partial^2 g / \partial z^2 = 0. \quad (3)$$

The orthogonal modes of the medium, in the $z > 0$ space, are denoted:

$$E_{j,m} = \psi_{j,m}(r,\theta) \exp[\pm i\beta_{j,m}z] \quad (4)$$

where the modal propagation constants can quite generally be written in the form

$$\beta_{j,m} = \left\{ 1 - \sum_p (a_p j^p + b_p m^p) \right\}^{1/2}, \quad (5)$$

(where we remind the reader of the length scaling just described).

With all sources located in the $z < 0$ half-space, the field in $z > 0$ is represented as a superposition of forward propagating modes

$$E(r,\theta,z) = \sum_{j,m} C_{j,m} \psi_{j,m}(r,\theta) \exp[i\beta_{j,m}z] \quad (6)$$

(We have neglected the radiation field in Eqn. (6). This will be an excellent approximation for highly multimoded, guiding media, and is an exact representation for infinite quadratic and similar profiles.)

The paraxial modes (which are characterized by having low values for m and j) have propagation constants close to 1. The approximation used in this chapter requires the exact modal propagation constants in expansion (6) to be replaced by

$$\tilde{\beta}_{j,m} = 1 - (1/2) \sum_p (a_p j^p + b_p m^p) \quad (7)$$

The approximate field is thus taken to be

$$\begin{aligned}
 \tilde{E}(r, \theta, z) &= \sum_{j,m} C_{j,m} \psi_{j,m}(r, \theta) \exp[i \tilde{\beta}_{j,m} z] \\
 &= \exp[iz] \tilde{g}(r, \theta, z).
 \end{aligned} \tag{8}$$

Now $\tilde{g}$ satisfies the equation

$$\partial^2 \tilde{g} / \partial r^2 + (1/r) \partial \tilde{g} / \partial r + (1/r^2) \partial^2 \tilde{g} / \partial \theta^2 - \alpha \tilde{g} + 2i \partial \tilde{g} / \partial z = 0 \tag{9}$$

A comparison of equations (3) and (9) shows that expressing the scalar field as a mode expansion and replacing the modal propagation constants, Eq. (5), by a first order binomial expansion, Eq. (7), is equivalent to writing the field in the form $E = \exp[iz] g$ and neglecting $\partial^2 g / \partial z^2$ as insignificant. Thus, the approximation involves representing the field as a rapidly varying function of z ($\exp[iz]$) modulated by a slowly varying envelope ($\tilde{g}$). This is the often-used parabolic (or slowly varying envelope) approximation of scalar wave theory¹⁰, and is the same approximation used to derive the Fresnel field in classical diffraction problems.

The relationship between the eigenfunction and eigenvalues of the Helmholtz and Parabolic equations was first discussed by Feit and Fleck.¹¹

5.3 DIFFRACTION IN A QUADRATIC-INDEX MEDIUM

With reference to Figure 1, consider a medium which has a refractive index that is quadratic in the $z > 0$ half-space,

$$n^2 = n_0^2 (1 - r^2/\rho^2), \quad z > 0. \quad (10)$$

where the constant ρ characterizes the degree of grading for the medium. Although the profile becomes unphysical for large r , ($n^2 \rightarrow -\infty$ as $r \rightarrow \infty$), this is of no consequence provided attention is focused on apertures of size much less than ρ , and provided the field is negligible everywhere except in the region $r/\rho < 1$. (It is also noted that like r and z , we have made ρ dimensionless by the scaling described in section 5.2).

The Laguerre-Gaussian modes of such a medium are¹²

$$\psi_{j,m} = r^j L_m^j(r^2/\rho) \exp[-r^2/2\rho] \begin{Bmatrix} \cos j\theta \\ \sin j\theta \end{Bmatrix} \quad (11)$$

$$j, m = 0, 1, 2 \dots$$

with modal propagation constants

$$\beta_{j,m} = \{ 1 - (2/\rho) (j + 2m + 1) \}^{1/2}. \quad (12)$$

(We, once again, remind the reader of the length scaling implicit in Eq.(12)).

For a general aperture (ξ in Figure 1) and arbitrary incident field, the scalar field in $z > 0$ is

$$\begin{aligned} E = & \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} A(j,m) r^j L_m^j(r^2/\rho) \exp[i\beta_{j,m}z - r^2/2\rho] \cos j\theta \\ & + \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} B(j,m) r^j L_m^j(r^2/\rho) \exp[i\beta_{j,m}z - r^2/2\rho] \sin j\theta. \end{aligned} \quad (13)$$

The modal amplitudes $A(j,m)$ and $B(j,m)$ are constants determined by the incident field and the shape of the aperture.

For an axisymmetric system, comprising a circular aperture with centre at $r = 0$ and radius a and a θ -independent incident field, the field in the aperture plane can be taken to be

$$E(r, z=0) = \begin{cases} \phi(r) & , \quad r < a \\ 0 & , \quad r > a \end{cases} \quad (14)$$

where we follow the usual approximation used in scalar diffraction theory by ignoring reflection effects and taking $\phi(r)$ to be what the incident field would be in the absence of the screen. Again, a is assumed to be dimensionless, having been scaled as described in section 5.2.

For an axisymmetric system Eq. (13) reduces to

$$E(r, z) = \sum_{m=0}^{\infty} A(o, m) L_m(r^2/\rho) \exp[-r^2/2\rho] \exp[i\beta_{o, m} z] \quad (15)$$

Orthogonality of Laguerre polynomials gives the modal amplitudes

$$A(o, m) = (2/\rho) \int_0^a s \phi(s) L_m(s^2/\rho) \exp[-s^2/2\rho] ds. \quad (16)$$

Substituting (16) into (15) gives the scalar field

$$E(r, z) = (2/\rho) \exp[-r^2/2\rho] \int_0^a s \phi(s) \exp[-s^2/2\rho] F(r, z, s) ds \quad (17)$$

where

$$F(r, z, s) = \sum_{m=0}^{\infty} L_m(r^2/\rho) L_m(s^2/\rho) \exp[i\beta_{o,m} z], \quad (18)$$

5.3.1 THE PARAXIAL FIELD

The paraxial field is obtained by replacing $\beta_{o,m}$ in Eq. (18) by

$$\tilde{\beta}_{o,m} = 1 - (1/\rho)(2m + 1) \quad . \quad (19)$$

When this is done and the summation is evaluated¹³ the paraxial field, $\tilde{E}$, is obtained:

$$\begin{aligned} \tilde{E}(r, z) = & (1/i\sigma(z)) \exp[i\{z + r^2/2\tau(z)\}] \\ & \times \int_0^a s \phi(s) \exp[is^2/2\tau(z)] J_0(rs/\sigma(z)) ds \end{aligned} \quad (20)$$

where

$$\begin{aligned} \sigma(z) &= \rho \sin(z/\rho) \\ \tau(z) &= \rho \tan(z/\rho) . \end{aligned} \quad (21)$$

The integral representation derived by Iga^{1,4-6} reduces to (20) for the special case of an axisymmetric system.

5.3.2 HOMOGENEOUS MEDIUM LIMIT

In the mathematical limit $\rho \rightarrow \infty$ the grading disappears and the profile becomes homogeneous. Noting that

$$\sigma(z) \rightarrow z$$

$$\tau(z) \rightarrow z$$

when this limit is applied, Eq. (20) becomes the correct result for diffraction in a homogeneous medium (see Appendix A, section 5.10.1).

5.4 PLANE WAVE NORMALLY INCIDENT WITH THIN LENS IN APERTURE

Eq. (20) is now applied to the specific case of plane waves normally incident on the aperture, with a thin lens inside the aperture. The lens introduces a phase transformation of the incident field so that to a good approximation the field in the aperture can be taken as ¹⁴

$$\phi(r) = \exp \left[-i r^2 / 2 f_H \right] \quad (22)$$

where f_H is the geometrical focal length of the thin lens when the $z > 0$ region is homogeneous, $n^2 = n_0^2$. A unit amplitude field has been assumed.

This situation actually includes a variety of cases. For example, taking the limit $f_H \rightarrow \infty$ is equivalent to having plane waves incident on a planar-faced graded-index medium. When f_H is finite, the above incident field (Eq. (22)) corresponds to having spherical waves of radius of curvature f_H striking the aperture. In this way, the discussion could have proceeded without any reference to thin lenses.

Substituting (22) into Eq. (20) gives the paraxial field

$$\begin{aligned} \tilde{E}(r,z) = & \left(a^2 / i \sigma(z) \right) \exp \left[i \left\{ z + r^2 / 2 \tau(z) \right\} \right] \\ & \times \int_0^1 t \exp \left[-i p(z) t^2 / 2 \right] J_0(q(r,z)t) dt \end{aligned} \quad (23)$$

where the functions $p(z)$ and $q(r,z)$ are defined by

$$p(z) = a^2 \left\{ \frac{1}{f_H} - \frac{1}{\tau(z)} \right\} \quad (24)$$

$$q(r,z) = a r / \sigma(z) \quad .$$

Following Born and Wolf¹⁵ this can be re-written in terms of Lommel functions, to give the intensity distribution behind the aperture as

$$\begin{aligned} \tilde{I}(r,\theta,z) &= \{f_H/(\sigma(z) - f_H \cos(z/\rho))\}^2 \\ &\times \{U_1^2(p(z), q(r,z)) + U_2^2(p(z), q(r,z))\} \end{aligned} \quad (25)$$

where U_1 and U_2 are Lommel functions, defined by¹⁵

$$U_n(p,q) = \sum_{s=0}^{\infty} (-1)^s (p/q)^{n+2s} J_{n+2s}(q) \quad . \quad (26)$$

An alternative expression to (25) involving the V_0 and V_1 Lommel functions could also be used, and Born and Wolf point out that this would be useful for numerical calculations with $q/p < 1$.

The homogeneous medium limit is discussed in Appendix A.

5.4.1 INTENSITY ALONG THE AXIS

From (23) the intensity along the z -axis is obtained:

$$\tilde{I}(r=0,z) = (a^4/4 \sigma^2(z)) \operatorname{sinc}^2(p(z)/4) \quad . \quad (27)$$

where the sinc function, as used here, is defined by $\text{sinc } x = (\sin x) / x$.

Focusing may be said to occur at locations given by the maxima of (27). Note that the sinc^2 part of Eqn. (27) varies considerably more rapidly with z , at optical frequencies, than the σ^2 part. (This point is obscured, somewhat, by the length scaling implicit in Eqn. (27).) Consequently, if there is a maximum at $z = z_m$ and z is constrained to a small neighbourhood of z_m then

$$\tilde{I} \approx (a^4 / 4 \sigma^2(z_m)) \text{sinc}^2(p(z)/4) \quad (28)$$

Thus we find that

$$z_m \approx \rho \tan^{-1}(f_H / \rho) \equiv z_{g.o,m} \quad (29)$$

The subscript, m , refers to the various branches of the $\tan^{-1}$ function. The equally spaced intensity maxima are assumed to be ordered in increasing values of m .

For the special case of no lens in the aperture ($f_H \rightarrow \infty$) Eq.(29) reduces to

$$z_m \approx \rho(2m + 1) \pi / 2 \equiv z_{g.o,m} \quad (30)$$

$$m = 0, 1, 2, \dots$$

It will be shown in section 5.6 that a geometrical optics analysis reveals that paraxial rays are focused at the locations specified by $z_{g.o,m}$. Consequently, these planes will be referred to as the geometrical focal planes. It is also interesting to note that taking the limit $\rho \rightarrow \infty$ in expression (29) (which corresponds to the gradedness of the medium disappearing) results in the infinity of equally spaced focal points being replaced by a single focal point, at $z = f_H$, as expected.

It must also be realized that for large enough values of z , higher order terms in the expansion of the modal propagation constants become significant and the analysis breaks down. Expression (27) is only valid for z restricted to a certain range (this is discussed quantitatively, later), and in practice only a finite number of focal planes will be found. This is responsible for the smearing of images transmitted through single GRIN rods, as discussed by various other authors^{2,3}.

It is instructive to change variable so that the origin of the axis is centred on the m^{th} geometrical focal plane:

$$z = z_{g.o,m} + z' . \quad (31)$$

Using z' as the new variable, the intensity along the axis can be written as

$$\tilde{I}(r=0, z') = \{ 4(\chi-1)/(\chi \sin^2(z'/\rho)) \} \quad (32)$$

$$\times \sin^2 \{ (a^2 \chi / 4 f_H) \tau(z') / (f_H + \tau(z')) \}$$

where χ is an important, new dimensionless parameter defined by

$$\chi = 1 + (f_H^2 / \rho^2) . \quad (33)$$

There are two distinct focusing effects in operation; the focusing action of the thin lens (characterized by f_H), and the self-focusing of the medium itself (characterized by ρ). The parameter, χ , combines together these two effects.

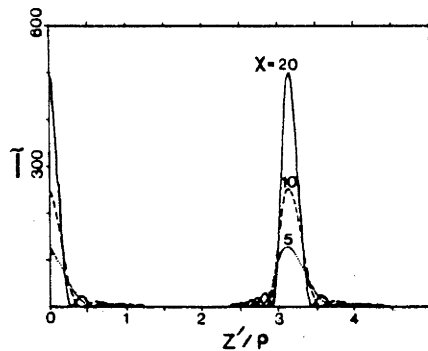


Fig. 2. Intensity along the axis. This is plotted for $(a^2/f_H) = 10$ and $\chi = 5, 10, 20$.

If f_H is finite, then clearly $\tilde{I}(r=0, z')$ has no symmetry and the intensity maxima are not located exactly at the geometrical focal planes. Thus we can speak of a "Focal Shift" just like the well-known focal shift of diffracted converging spherical waves in homogeneous media^{16,17}. In Figure 2 we plot the intensity against z'/ρ with typical parameter values, for various values of χ . The geometrical focal planes are located at $z'/\rho = 0, \pm \pi, \pm 2\pi \dots$ and the plots clearly exhibit the asymmetry and the focal shift (each intensity maximum is found at a point between the aperture and its associated geometrical focal plane). As χ increases, the asymmetry and focal shift both diminish, and the intensity peaks narrow and grow in height as light becomes more tightly concentrated at the focus.

Taking the limit $f_H \rightarrow \infty$ of expression (32) gives the intensity along the axis without a lens in the aperture (corresponding to plane waves normally incident onto the flat endface of the medium):

$$\tilde{I}(r=0, z') = 4 \sin^2\{a^2 \tan(z'/\rho)/4\rho\} / \sin^2(z'/\rho) \quad . \quad (34)$$

In this case, the intensity is an even function of z' and so the intensity along the axis is symmetrical with respect to the geometrical focal planes. Furthermore, an expansion of (34) to second-order in z'/ρ gives

$$\tilde{I}(r=0, z') \approx (a^4/4\rho^2) \{1 + (1 - a^4/48\rho^2)(z'/\rho)^2\} \quad (35)$$

$$z'/\rho \ll 1$$

Clearly, the intensity on the geometrical focal planes is a maximum if $(a^2/\rho) > 4/3$ and a local minimum if $(a^2/\rho) < 4/3$. In Figure 3 we plot intensity along the axis against (z'/ρ) for various different values of (a^2/ρ) . When $(a^2/\rho) \gg 4/3$, the intensity maxima are observed to be narrow and centred exactly on the geometrical focal planes (i.e. there is no focal shift). As (a^2/ρ) decreases, the peaks broaden and diminish in height until the critical value, $4/3$, is passed. When this happens, each intensity peak splits into two secondary peaks (symmetrically placed on either side of their associated geometrical focal planes), and the intensity is actually a local minimum on each geometrical focal plane. This splitting of the intensity maxima is an unusual effect that has no analog with focusing in homogeneous media.

Similar behaviour can be observed for the case when there is a lens in the aperture, but the effect is far less dramatic because it occurs when the asymmetry of the peaks is high, and this tends to mask the effect. Figure 4 illustrates this.

Even with a lens in the aperture, for a wide range of parameter values the asymmetry is slight and we can speak of the "half-width" of the intensity distribution along the axis, Ω_p . We define Ω_p as the value of z' for which the first zero of Eq. (32) is obtained. After some simplifying approximations we obtain

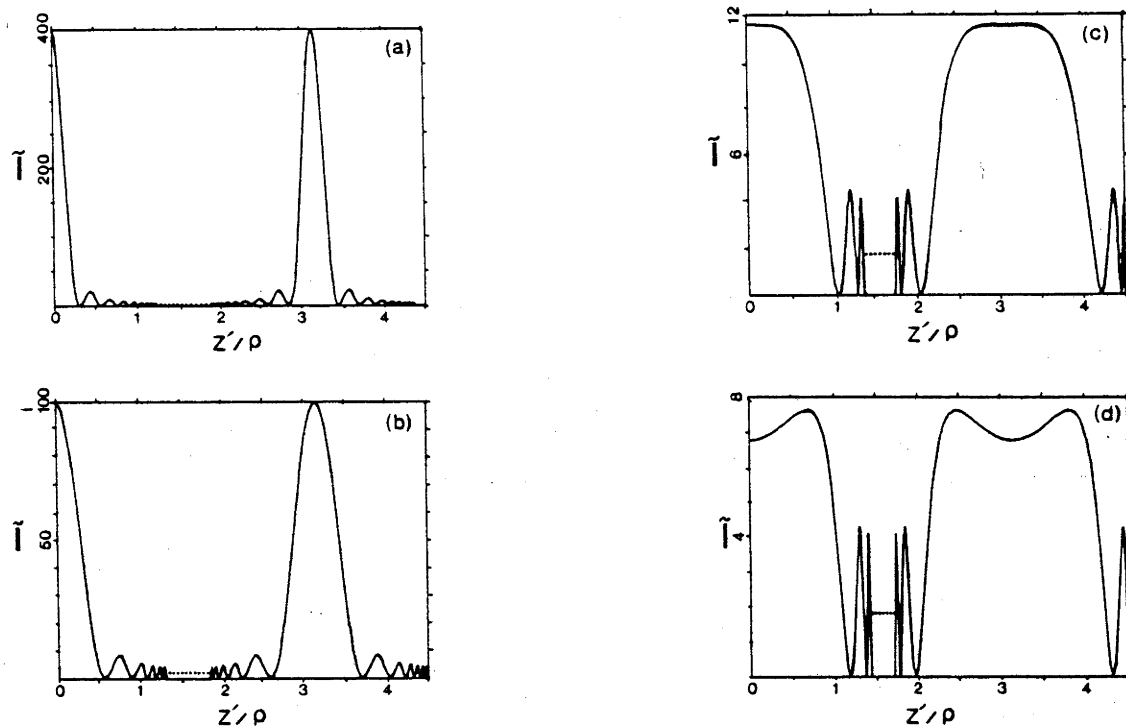


Fig. 3. Intensity along the axis (without a lens in the aperture). This is shown for various values of (a^2/ρ) . (a) $(a^2/\rho) = 4 \times 10$; (b) 4×5 ; (c) $4 \times \sqrt{3}$ (the critical value at which the splitting of the intensity maxima occurs); (d) 4×1.3 . (The dotted regions denote rapid oscillation of the intensity that cannot be reliably drawn.)

$$\Omega_p \approx 4\pi f_H^2 / (a^2 \chi) \quad . \quad (36)$$

Taking the limit $\rho \rightarrow \infty$ of expression (36) gives the "half-width" in the homogeneous case,

$$\Omega_H \approx 4\pi f_H^2 / a^2 \quad , \quad (37)$$

which shows that the effect of adding the quadratic gradient is simply to scale the intensity width by the factor $(1/\chi)$.

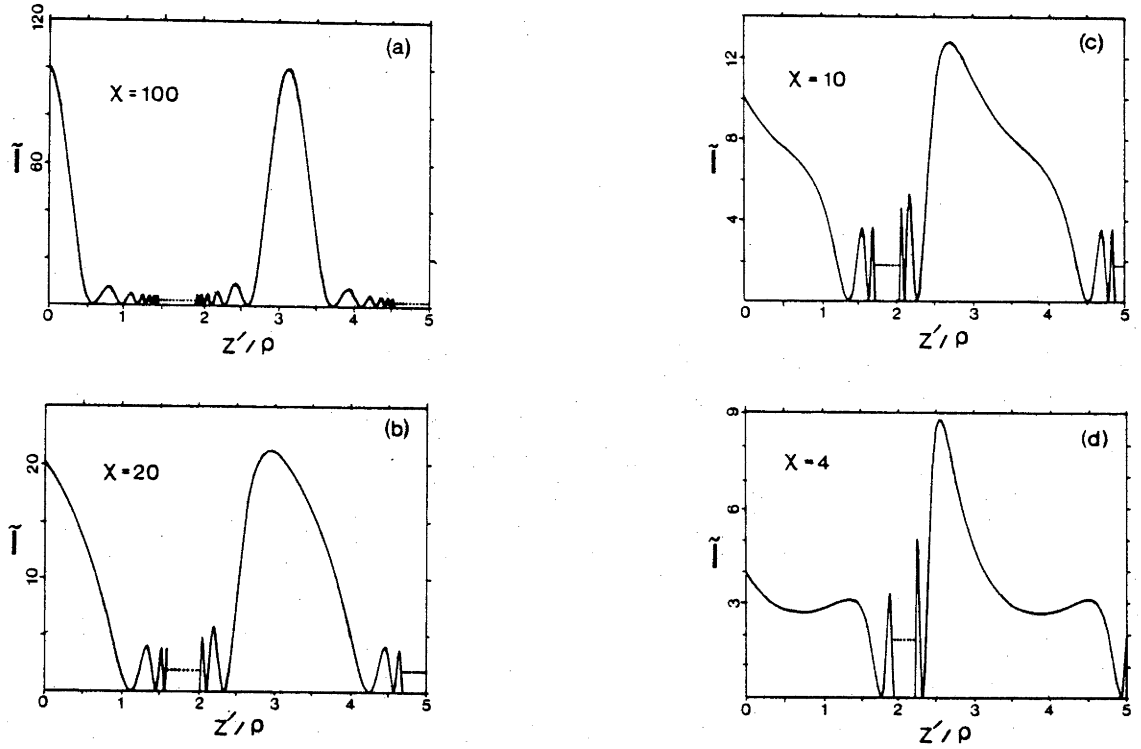


Fig. 4. Intensity along the axis. This is plotted for $(a^2/f_H) = 2$ and $\chi =$ (a) 100, (b) 20, (c) 10, and (d) 4. (The dotted regions denote rapid oscillation of the intensity that cannot be reliably drawn.)

5.4.2 INTENSITY IN THE GEOMETRICAL FOCAL PLANE

From Eq. (23) the intensity distribution in the geometrical focal plane is found to be

$$\tilde{I}(r, z=z_{g.o,m}) = (a^4 \chi / 4 f_H^2) \{ 2J_1(ar\sqrt{\chi}/f_H) / (ar\sqrt{\chi}/f_H) \}^2. \quad (38)$$

Defining the "half-width" of the intensity distribution, ω_p , to be the value of r for which the first zero of (38) is obtained, leads to the result

$$\omega_p = 1.22\pi f_H / (a\sqrt{\chi}). \quad (39)$$

Again we take the homogeneous medium limit of (39) to obtain

$$\omega_H = 1.22\pi f_H / a \quad ; \quad (40)$$

the well-known Airy radius for diffraction in a homogeneous medium.

Thus, the intensity distribution in the geometrical focal plane is simply a scaled version of the homogeneous case. The magnitude of the distribution is scaled by χ , and the width is scaled by $1 / \sqrt{\chi}$.

5.4.3 SYMMETRY OF THE FOCUSED FIELD

It is convenient to change variable from z to z' (see Eq. (31)) in expressions (23) and (24) to give the intensity distribution behind the apertured lens as

$$\tilde{I} = \{(\chi-1) / (\chi \sin^2(z'/\rho))\} \{U_1^2(u,v) + U_2^2(u,v)\} \quad (41)$$

where

$$u = (a^2\chi/f_H) \{ \tau(z') / (f_H + \tau(z')) \} \quad (42)$$

$$v = \{a\sqrt{\chi} / \cos(z'/\rho)\} \{1 / (f_H + \tau(z'))\} .$$

With the new z' variable the geometrical optics foci are located at $(z'/\rho) = 0, \pm\pi, \pm2\pi, \dots$. From (41), (42) and the symmetry properties of the Lommel functions:

$$U_1(-u,v) = -U_1(u,v) \quad (43)$$

$$U_2(-u,v) = U_2(u,v)$$

we find that if there is no lens in the aperture ($f_H \rightarrow \infty$) then the intensity distribution exhibits reflection symmetry about the geometric focal planes:

$$\tilde{I}(r, z') = \tilde{I}(r, -z') \quad . \quad (44)$$

However, if there is a lens present, then due to the term

$$\{1/(f_H + \tau(z'))\}$$

appearing in u and v , such symmetry no longer exists.

The fact that reflection symmetry exists when there is no lens present is not a trivial result. Geometrical optics considerations would suggest that this should be the case, but suggests the same for focusing of plane waves by a lens in a homogeneous medium. In Appendix A (section 5.10.1) we point out that this is not the case when we discuss the well-known focal shift effect.

5.5 VALIDITY OF THE PARAXIAL APPROXIMATION

5.5.1 RESTRICTIONS ON PARAMETERS

The approximation developed in this chapter, and detailed formally in section 5.2, requires the replacement of the exact modal propagation constants, Eq. (12), by the first order binomial expansion, Eq. (19), in the modal representation of the field. Clearly, only those modes with

$$m \ll \rho/4 \quad (45)$$

are well approximated by such a replacement. Evanescent modes, in particular, are very poorly described. Thus, the approximation will be accurate only if the spectrum of modal amplitudes, excited by the incident field, is essentially cut-off at a value of m much less than $(\rho/4)$. This requirement leads to restrictions on the magnitudes of the parameters, a and ρ .

Consider the simplest case: plane waves normally incident on an aperture without a lens. The spread in the spectrum of modal amplitudes, $A(o,m)$, is related to the width of the aperture, a , but in a far more complicated way than in the well-known Fourier case. In order to calculate the modal amplitudes explicitly, we model the step-function aperture field distribution

$$E(r,o) = \begin{cases} 1, & r < a \\ 0, & r > a \end{cases} \quad (46)$$

by a Gaussian function

$$E(r,o) = \exp [-r^2/2a^2] \quad . \quad (47)$$

With aperture field given by (47), the magnitude of the modal amplitudes is found to be

$$|A(o,m)| = A(0,0) \exp [-C_1 m] \quad (48)$$

where

$$C_1 = \begin{cases} 2 \tanh^{-1} (a^2/\rho) , & (a^2/\rho) \leq 1 \\ 2 \coth^{-1} (a^2/\rho) , & (a^2/\rho) > 1 \end{cases} \quad (49)$$

We can define the width of the spectrum, ΔA , as:

$$\Delta A = (1/C_1) = \begin{cases} 1/[2 \tanh^{-1} (a^2/\rho)] , & (a^2/\rho) \leq 1 \\ 1/[2 \coth^{-1} (a^2/\rho)] , & (a^2/\rho) > 1 \end{cases} \quad (50)$$

The relationship between ΔA and a (Eq.(50)) is an expression of the "Uncertainty Principle" for Laguerre-Gaussian expansions. Figure 5 shows a plot of ΔA against (a^2/ρ) . The minimum occurs when the Gaussian aperture field distribution exactly matches the Gaussian fundamental mode of the medium, so that a single mode is excited.

Clearly, the paraxial approximation will be accurate provided

$$\Delta A \ll \rho/4 \quad , \quad (51)$$

and from Figure (5) we see that this will be true if

$$\tanh(2/\rho) \ll a^2/\rho \ll 1/\tanh(2/\rho) \quad . \quad (52)$$

In addition to this, the quadratic profile (Eq.(10)) leads to unphysical results unless

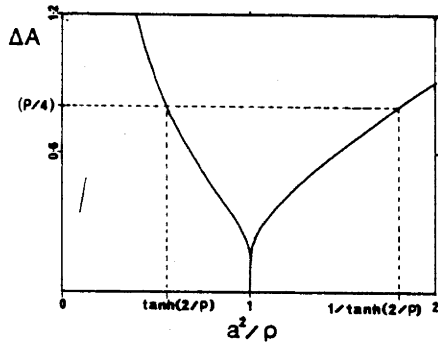


Fig. 5. Uncertainty relation for Laguerre-Gaussian expansions.

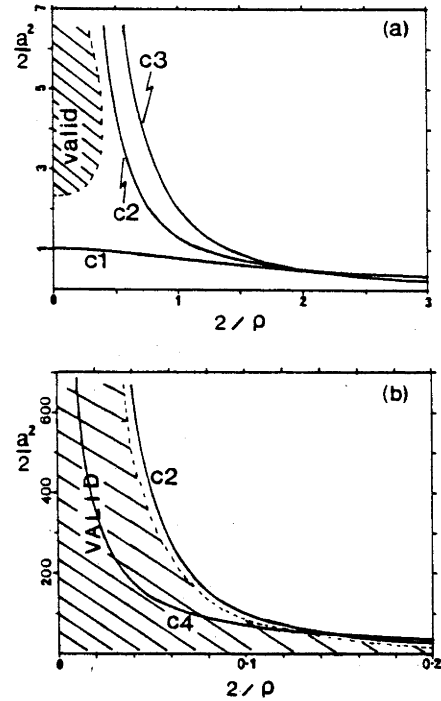


Fig. 6. (a) Region of validity of the analysis. We have plotted the curves $C1, \tanh(2/\rho)/(2/\rho)$; $C2, 1/[(2/\rho) \tanh(2/\rho)]$; $C3, 2/(2/\rho)^2$. The shaded region represents the set of points $(2/\rho, a^2/2)$ for which the analysis is valid. (b) Blowup of (a) in the region $0 < (2/\rho) < 0.2$. The curves are $C4, 4\sqrt{3}/(2/\rho)$, and $C2, 1/[(2/\rho) \tanh(2/\rho)]$. [On this scale the curve $\tanh(2/\rho)/(2/\rho)$ is indistinguishable from the horizontal axis.]

$$a^2/\rho^2 \ll 1$$

(53)

We combine (52) and (53) to obtain a convenient expression for the conditions imposed on the parameters, which ensures the validity of the analysis:

$$\frac{\tanh(2/\rho)}{(2/\rho)} \ll a^2/2 \ll \min \left\{ \begin{array}{l} \{ (2/\rho) \tanh(2/\rho) \}^{-1} \\ 2 / (2/\rho)^2 \end{array} \right\}.$$

(54)

(the reader is once again reminded of the wavelength normalization of all length quantities, that was described in section 5.2. Thus both a and ρ are understood to be dimensionless).

Figure 6(a) illustrates the region of validity of the analysis (note that the dotted lines in Figure (6) give a schematic guide only to the validity region). The shaded region represents the set of points $(2/\rho, a^2/2)$ which satisfy (54). It is clear from the diagram that only when $(2/\rho) \lesssim 0.5$ can the required conditions be satisfied.

It is interesting to note that taking the homogeneous medium limit (i.e., $\rho \rightarrow \infty$) of (54) reduces the validity condition to the simple restriction

$$a^2 \gg 1, \quad (55)$$

which is an often used assumption in scalar diffraction theory.

Finally, we investigate the parameters for which the splitting of the central maxima occurs. To this end, in Figure 6(b) we have shown a blown-up version of Figure 6(a) and have added the curve $4\sqrt{3} / (2/\rho)$. If the parameters of the system are such that the point $(2/\rho, a^2/2)$ lies above this curve then the intensity maxima are located exactly on the geometrical focal planes. However, if the parameters are such that $(2/\rho, a^2/2)$ lies beneath this curve, then the intensity maxima have split and the intensity on each geometrical focal plane is a local minimum. Thus we have shown that the splitting occurs for parameter values well within the region of validity of the analysis, and the splitting phenomenon is expected to be a real, physical, effect.

We return, in section 5.6, to the physics behind the restrictions embodied in Eq. (54).

5.5.2 RESTRICTIONS ON Z

There are also restrictions placed on the observation point. When z becomes too large, the higher-order terms neglected from the first-order binomial expansion of the modal propagation constants (Eq. (7)) may become significant. For an axisymmetric quadratic-index medium, replacing the exact modal propagation constants, $\beta_{o,m}$ (Eq. (12)) by $\tilde{\beta}_{o,m}$ (Eq. (19)) will lead to accurate results only if

$$\{(2m+1)^2 / 2\rho^2\} z \ll \pi \quad . \quad (56)$$

The severest restriction is due to those modes at the far end of the modal amplitude spectrum. For the simplest case; plane waves normally incident on an aperture without a lens, the "worst-case" modes have $m \sim \Delta A$ where ΔA is given by Eq. (50). We thus obtain the following condition for validity of the paraxial approximation:

$$z \ll 2\pi\rho^2 T^2(a^2/\rho) \quad (57)$$

where

$$T(a^2/\rho) = \begin{cases} \tanh^{-1}(a^2/\rho) / [1 + \tanh^{-1}(a^2/\rho)], & a^2/\rho < 1 \\ \coth^{-1}(a^2/\rho) / [1 + \coth^{-1}(a^2/\rho)], & a^2/\rho > 1 \end{cases} \quad (58)$$

In the limit of the gradedness of the medium disappearing, Eqn. (57) reduces to the simple condition

$$z \ll 2a^4 \quad . \quad (59)$$

5.6 GEOMETRIC OPTICS

It is interesting to compare some of the results derived in previous sections with the predictions of geometrical optics. Note that the meridional rays in quadratic-index media are described by simple sinusoidal trajectories.¹²

A ray normally incident on the aperture, a radial distance r_i from the centre, is easily shown to cross the z -axis periodically at

$$z = \{ \rho f_H Q^2 / (f_H^2 Q^2 + r_i^2) \}^{1/2} \tan^{-1} \{ (f_H^2 Q^2 + r_i^2)^{1/2} / \rho Q \} \quad (60)$$

where

$$Q = \{ 1 - r_i^2 / \rho^2 \}^{1/2} = n(r_i) / n_0 \quad . \quad (61)$$

Taking the limit $f_H \rightarrow \infty$ of expression (60) gives the z -axis crossings when there is no lens in the aperture,

$$z = \{ \rho (2m + 1) \pi / 2 \} Q \quad . \quad (62)$$

Equations (60) and (62) show that rays normally incident at different radial distances from the centre of the aperture are brought to different foci (illustrating the well-known fact that a quadratic-index medium is not ideally focusing¹⁸). However, if only paraxial rays with

$$r_i \ll \rho \quad r_i \ll f_H \quad (63)$$

are considered, then expressions (60) and (62) reduce to $z_{g.o,m}$ (Eqs. (29)

and (30)). Thus it is appropriate to speak of the focusing of paraxial rays in a quadratic-index medium (this point is discussed in detail by Mikaelian¹⁸), and we have identified these foci as the locations $z_{g.o,m}$ derived in section 5.4.

Ray optics provides a simple, physical explanation for the condition (Eq. (54)) imposed on the parameters a and ρ . We first note that we can approximate (54) to

$$1 \ll a^2/2 \ll 1/(2/\rho)^2, \quad (64)$$

by using the previously mentioned fact that $(2/\rho) \lesssim 0.5$. The breakdown of the analysis when (64) is not satisfied, is due to the growing importance of non-paraxial rays as a^2 approaches either limit.

Suppose we make a so small that $1 \ll a^2$ is no longer satisfied. Due to diffraction effects at the aperture, the incident rays can no longer be assumed to be normal to the aperture plane, but are spread by the angle

$$\theta_{\text{diff}} \sim 1.22\pi/a. \quad (65)$$

Although this spread is small if (64) is satisfied, it leads to a significant amount of incident power being coupled into non-paraxial rays as a^2 approaches the lower limit. Fig. 7a illustrates this effect.

At the other extreme, if a is so large that the upper limit of (64) is approached, then diffraction effects at the aperture are negligible, but the aperture is accepting rays which are incident relatively large distances from the centre. From (63) we see that such rays are non-paraxial, as illustrated schematically in Fig. 7b.

5.7 OBLIQUE INCIDENCE

We have discussed the focusing of normally incident plane waves, up to this point, but will now consider plane waves incident at an angle, δ , to the z axis (see Figure 8). This situation is most important when considering the angular sensitivity of photoreceptor excitation, as discussed later this section.

The x - y co-ordinates in the aperture plane are oriented so that the incident wave-vector lies in the x - z plane. After phase transformation by the thin lens, the field in the aperture plane can be taken to be

$$E(r, \theta, z=0) = \begin{cases} \phi(r, \theta), & r < a \\ 0, & r > a \end{cases} \quad (66)$$

where

$$\phi(r, \theta) = \exp[i(n_1/n_0)(\sin\delta)r\cos\theta] \exp[-ir^2/2f_H] \quad (67)$$

The incident field is no longer axisymmetric so expression (20) cannot be used. The mathematical details of the derivation of the paraxial field are given in Appendix B (section 5.10.2). Here we shall only state the result (Eqs. (B12) and (B13)) :

$$\begin{aligned} \tilde{E}(r, \theta, z) = & (a^2/i\sigma(z)) \exp[i\{z + r^2/2\tau(z)\}] \\ & \times \int_0^1 s \exp[-ip(z)s^2/2] J_0(Kq(r, z)s) ds \quad , \quad (68) \end{aligned}$$

where

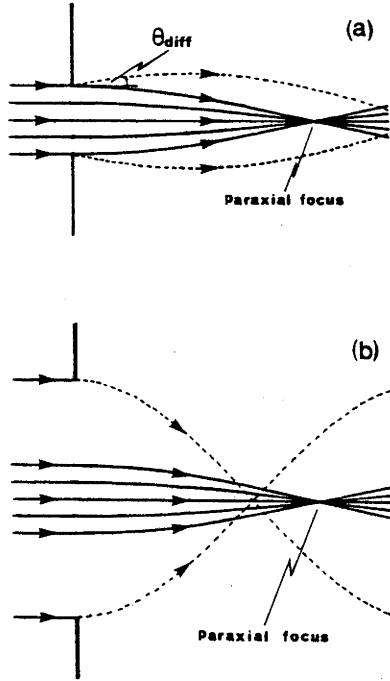


Fig. 7. (a) Schematic of the formation of nonparaxial rays due to diffraction effects at the aperture when the aperture radius is too small. The dotted rays are the nonparaxial ones. (b) Nonparaxial rays (dotted curves) entering when the aperture is too large.

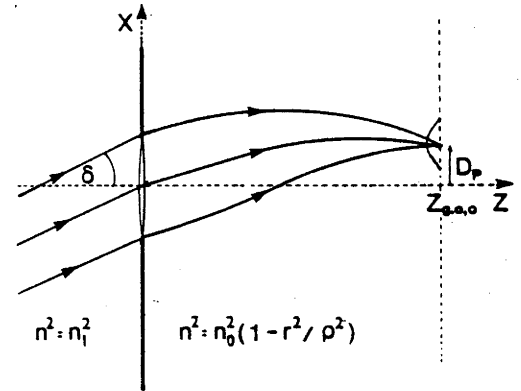


Fig. 8. Plane waves incident at an angle δ to the z axis. Also shown is the first geometrical focal plane $z_{g.o.o}$ and the translation of the focused field D_p .

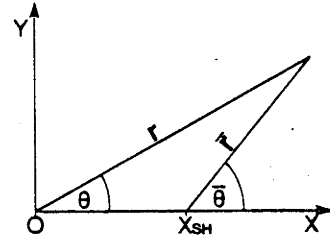


Fig. 9. New coordinate system $(\bar{r}, \bar{\theta})$. The new coordinate system is obtained by translating the origin of the old system, a distance $X_{sh} = (n_1/n_0)(\sin \delta) \sigma(z_a)$ along the X axis, where we are interested in the field on the plane $z = z_a$.

$$\kappa = \left\{ 1 - 2(n_1/n_0)(\sin \delta) \sigma(z) \cos \theta / r + (n_1/n_0)^2 (\sin \delta)^2 \sigma^2(z) / r^2 \right\}^{1/2} \quad (69)$$

and $\sigma(z)$, $\tau(z)$, $p(z)$, $q(r, z)$ are defined in Eqs. (21) and (24) respectively.

It is informative to change co-ordinates in the aperture plane from (r, θ) to $(\bar{r}, \bar{\theta})$ where the new co-ordinate system is obtained from the old one as follows: if we wish to evaluate the field on the plane $z = z_a$, then we translate the origin of the old $r-\theta$ co-ordinate system by

an amount $(n_1/n_0)(\sin\delta)\sigma(z_a)$ along the x-axis, as illustrated in Figure 9. Thus, the amount of translation depends on the particular z-plane on which the field is to be determined.

Referring to Figure 9, it follows that

$$K = \bar{r}/r \quad (70)$$

and hence

$$K q(r, z) = q(\bar{r}, z) \quad (71)$$

The intensity distribution behind the aperture can be written in terms of Lommel functions in the same way as for the normally incident case (Eq. (25)) :

$$\tilde{I}(\bar{r}, z) = \{f_H / (\sigma(z) - f_H \cos(z/\rho))\}^2 \quad (72)$$

$$\times \{U_1^2(p(z), q(\bar{r}, z)) + U_2^2(p(z), q(\bar{r}, z))\} \quad .$$

By comparing the intensity distribution for oblique incidence (Eq.(72)), with the analogous expression for normal incidence (Eq. (25)), we see that the two are very simply related: the intensity distribution on the plane $z = z_a$ for the case of oblique incidence is exactly the same as the intensity distribution found for normal incidence, except that it is translated along the x-axis by an amount $(n_1/n_0)(\sin\delta)\sigma(z_a)$.

Of special interest to photoreceptor optics is the case when z_a is one of the geometrical focal planes. This is because an important consideration in photoreceptor systems is the angular sensitivity, usually

defined as the amount of power absorbed by the photoreceptor as the angle at which light is incident on the lens-photoreceptor system is varied. Pask and Barrell⁹ have carried out a detailed analysis of the optical factors influencing angular sensitivity, using a simplified model involving a homogeneous medium separating the lens from the photoreceptor. The work presented in this section illustrates how the work of Pask and Barrell can be easily generalized to cover the more complex model involving a radially graded medium separating the lens from the photoreceptor. As the angle of incidence, δ , increases, the Airy pattern intensity distribution translates across the geometrical focal plane (and thus across the surface of the photoreceptor). For small angles of incidence ($\delta \ll 1$) the Airy pattern shifts a distance

$$D_p \approx (n_1/n_o) f_H \delta / \sqrt{\chi} \quad . \quad (73)$$

Using Eq. (73) the angular sensitivity can be calculated using the method described by Pask and Barrell.

Finally we note that in the homogeneous limit, (73) reduces to the well known result⁹

$$D_H \approx (n_1/n_o) f_H \delta \quad . \quad (74)$$

5.8 COMPARISON OF FOCUSING IN QUADRATIC AND HOMOGENEOUS MEDIA

Here we summarize the special features of focusing in quadratic index media, and compare with focusing in uniform media.

The half-widths of the intensity distribution along the z-axis and in the geometrical focal plane are given by Eqs. (36) and (39), respectively, for the quadratic medium, and by Eqs. (37) and (40) for the homogeneous case. We see that:

$$\Omega_H / \Omega_P = \chi > 1$$

(75)

$$\omega_H / \omega_P = \sqrt{\chi} > 1 .$$

The effect of the quadratic grading is to reduce the width of the focused field, in both orthogonal directions. Also, the intensity at the geometrical foci is greater for the quadratic case:

$$\tilde{I}_P(r=0, z'=0) / \tilde{I}_H(r=0, z'=0) = \chi > 1 . \quad (76)$$

Thus, adding the transverse gradient to a homogeneous medium results in a tighter concentration of light in both orthogonal directions.

It is interesting to note that the total power contained within the central spot of the Airy pattern is the same for both media. This follows after noting that the total power contained within a circle of radius r_0 in the geometrical focal plane for the homogeneous medium is¹⁵

$$\tilde{P}_H(r_0) = \pi a^2 \{ 1 - J_0^2(ar_0/f_H) - J_1^2(ar_0/f_H) \} , \quad (77)$$

and for the quadratic medium it is

$$\tilde{P}_P(r_0) = \pi a^2 \{ 1 - J_0^2(ar_0\sqrt{\chi}/f_H) - J_1^2(ar_0\sqrt{\chi}/f_H) \} . \quad (78)$$

Setting r_0 equal to ω_H in (77) and ω_P in (78) gives the desired result.

Accurate analyses of the intensity distribution in the focal region of a converging spherical wave diffracted by a circular aperture in

a homogeneous medium^{16,17}, show that the intensity is asymmetrical with respect to the geometrical focal plane. In fact, the intensity maximum occurs at some point between the aperture and geometrical optics focus (this point is discussed further in Appendix A).

The results in this chapter have shown that in the case of a quadratic-index medium, if the focusing of the incident plane waves is due entirely to the self-focusing of the medium itself (i.e. no lens in the aperture), then the intensity distribution exhibits reflection symmetry with respect to the geometrical focal planes. Furthermore, if $(a^2/\rho) > 4/3$ then the intensity maxima lie exactly on the geometrical focal planes (i.e. no focal shift), but if $(a^2/\rho) < 4/3$ then the intensity maxima have split into two secondary maxima lying, symmetrically, on either side of the associated geometrical focal planes, with a local minimum on the geometrical focal planes.

The presence of a lens destroys this symmetry, and the intensity maxima are always located between the lens and the associated geometrical focal planes (i.e. there is a focal shift).

5.9 CONCLUSION

We have presented a detailed study of focusing in quadratic-index media, and emphasized the fundamental differences with the focused field of a diffraction-limited lens in a homogeneous medium.

In the next chapter, we extend the theory given here to cover the case of slowly tapered quadratic-index media, in order to point out the most fundamental changes resulting from the tapering.

5.10.1 APPENDIX A: HOMOGENEOUS MEDIUM LIMIT

Taking the limit $\rho \rightarrow \infty$ of expression (20) gives the paraxial field for the homogeneous medium, $n^2 = n_0^2$:

$$\begin{aligned} \tilde{E}(r,z) = & (1/iz) \exp[i\{z + r^2/2z\}] \\ & \times \int_0^a s \phi(s) \exp[is^2/2z] J_0(rs/z) ds. \end{aligned} \quad (A1)$$

This same expression could have been derived directly for the homogeneous medium by expressing the field as an expansion over its plane wave modes (the well-known Angular Spectrum representation of the field) and linearizing the modal propagation constants in the manner described in section 5.2.

As pointed out by Iga et. al⁶, expression (A1) is a well-known Fresnel-Kirchhoff integral, and thus Eq. (20) exhibits the correct limiting behaviour as $\rho \rightarrow \infty$.

For the case of plane waves normally incident on a thin lens in the aperture, the field in the aperture plane is given by (22). Substituting (22) into (A1) and changing variable z to z' where

$$z = f_H + z', \quad (A2)$$

(i.e. the origin of the z' -axis is at the geometrical optics focus) gives the result:

$$\tilde{E}(r, z') = \{a^2/i(f_H + z')\} \exp[i\{z' + r^2/2(f_H + z')\}] \quad (A3)$$

$$\times \int_0^1 s \exp[-ia^2 s^2 z'/2f_H(f_H + z')] J_0(ars/(f_H + z')) ds$$

Eq. (A3) is the same result as calculated by Erkkila and Rogers¹⁶ for the light distribution near focus of a converging spherical wave diffracted by a circular aperture. Erkkila and Rogers derived the result using an improved approximation over the classical treatments¹⁵, and they point out that the intensity predicted by (A3) exhibits reflection asymmetry with respect to the geometrical focal plane, and leads to an intensity maximum located between the aperture and geometrical focal plane. Li and Wolf¹⁷ first used the term "focal shift" to refer to this effect.

5.10.2 APPENDIX B: OFF AXIS INCIDENCE

Since the incident field is no longer axisymmetric, we refer to the general expression for the diffracted field, Eq. (13). Because $E(r, \theta, z=0)$ is an even function of θ , we must have

$$B(j, m) = 0 ; \quad j, m = 0, 1, 2, \dots \quad (B1)$$

Thus the scalar field in the $z > 0$ half-space is

$$E(r, \theta, z) = \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} A(j, m) r^j L_m^j(r^2/\rho) \exp[i\beta_{j,m} z - r^2/2\rho] \cos j\theta. \quad (B2)$$

By application of orthogonality of cosine functions followed by orthogonality of associated Laguerre functions we obtain the modal amplitudes:

$$A(j, m) = (4/\pi\rho) (1 - \delta_{j,0}/2) (1/\rho^j) m!/(m+j)! \quad (B3)$$

$$\times \int_{t=0}^a \int_{v=0}^{\pi} \exp[\eta(t, v)] t^{j+1} L_m^j(t^2/\rho) \cos jv \, dt \, dv$$

where

$$\eta(t, v) = i(n_1/n_0)(\sin\delta)t\cos v - (it^2/2f_H) - (t^2/2\rho). \quad (B4)$$

The paraxial field is obtained by replacing $\beta_{j,m}$ in (B2) by

$$\tilde{\beta}_{j,m} = 1 - (1/\rho) (j + 2m + 1). \quad (B5)$$

This leads to

$$\tilde{E}(r, \theta, z) = (4/\pi\rho) \gamma(z) \exp[iz - r^2/2\rho] \quad (B6)$$

$$\times \int_{t=0}^a \int_{v=0}^{\pi} t \exp[\eta(t, v)] G(r, \theta, z; t, v) \, dt \, dv,$$

where

$$G(r, \theta, z; t, v) = \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} (1 - \delta_{j,0}/2) (m!/(m+j)!) \cos j\theta \quad (B7)$$

$$\times \cos jv \cdot \{r\gamma(z)t/\rho\}^j \{\gamma^2(z)\}^m L_m^j(r^2/\rho) L_m^j(t^2/\rho),$$

and

$$\gamma(z) = \exp[-i z / \rho] . \quad (B8)$$

The m -summation is found in Erdelyi¹³ and is carried out immediately. Next, the v -integration follows from the familiar integral representation of integer-order Bessel functions¹³:

$$J_j(x) = (1/\pi i^j) \int_0^\pi \exp[ix \cos v] \cos jv \, dv .$$

After simplification, the paraxial field takes the form:

$$\begin{aligned} \tilde{E}(r, \theta, z) = & (2a^2/i\sigma(z)) \exp[i\{z + r^2/2\tau(z)\}] \\ & \times \int_0^1 t \exp[-ip(z)t^2/2] H(r, \theta, z; t) \, dt \end{aligned} \quad (B9)$$

where $\sigma(z)$, $\tau(z)$ and $p(z)$ are defined in Eqs. (21) and (24) respectively, and

$$H(r, \theta, z; t) = \sum_{j=0}^{\infty} (1 - \delta_{j,0}/2) J_j(\omega_1) J_j(\omega_2) \cos j\theta \quad (B10)$$

$$\omega_1 = ar t / \sigma(z)$$

$$\omega_2 = a(n_1/n_0)(\sin\delta)t.$$

The j -summation, defining $H(r, \theta, z; t)$, is given by a well-known addition theorem for Bessel functions¹³,

$$H(r, \theta, z; t) = (1/2) J_0(\{\omega_1^2 + \omega_2^2 - 2\omega_1\omega_2\cos\theta\}^{1/2}). \quad (B11)$$

Hence, we obtain the paraxial field in its final form:

$$\begin{aligned} \tilde{E}(r, \theta, z) = & (a^2/i\sigma(z)) \exp[i\{z + r^2/2\tau(z)\}] \\ & \times \int_0^1 t \exp[-ip(z)t^2/2] J_0(K q(r, z) t) dt \end{aligned} \quad (B12)$$

where $q(r, z)$ is defined in Eq. (24), and

$$K = \left\{ 1 - 2(n_1/n_0)(\sin\delta)\sigma(z)\cos\theta/r + (n_1/n_0)^2(\sin\delta)^2\sigma^2(z)/r^2 \right\}^{1/2} \quad (B13)$$

5.11 REFERENCES

1. K. Iga, S. Hata, Y. Kato and H. Fukuyo, "Image Transmission by an Optical System with a Lens-like Medium," Japan. J. Appl. Phys. 13, 79 (1974).
2. A. Yariv, "On Transmission and Recovery of Three-Dimensional Image Information in Optical Waveguides," J. Opt. Soc. Am. 66, 301 (1976).
3. A.A. Friesem, U. Levy and Y. Silberberg, "Parallel Transmission of Images Through Single Optical Fibres," Proc. IEEE 71, 208 (1983).

4. K. Iga, "Theory for Gradient-Index Imaging," Appl. Opt. 19, 1039 (1980).
5. There is an error in the integral transform expression for the field given in reference 4. The correct expression can be found in reference 6.
6. K. Iga, Y. Kokubun and M. Oikawa, "Fundamentals of Micro-Optics, Distributed-Index, Microlens, and Stacked Planar Optics," (Academic, New York, 1984).
7. M.S. Sodha and A.K. Ghatak, "Inhomogeneous Optical Waveguides," (Plenum, New York, 1977).
8. C. Pask and K.F. Barrell, "Photoreceptor Optics I: Introduction to Formalism and Excitation in a Lens-Photoreceptor System," Biol. Cybernetics 36, 1 (1980).
9. C. Pask and K.F. Barrell, "Photoreceptor Optics II: Application to Angular Sensitivity and Other Properties of Lens-Photoreceptor System," Biol. Cybernetics 36, 9 (1980).
10. F.D. Tappert, "The Parabolic Approximation Method," Wave Propagation and Underwater Acoustics, J.B. Keller and J.S. Papadakis, eds. (Springer-Verlag, Berlin, 1977).

11. M.D. Feit and J.A. Fleck, "Calculation of Dispersion in Graded-Index Multimode Fibres by a Propagating-Beam Method," Appl. Opt. 18, 2843 (1979).
12. A.W. Snyder and J.D. Love, "Optical Waveguide Theory," (Chapman and Hall, London, 1983).
13. A. Erdelyi (Ed.), "Higher Transcendental Functions Vol. II," Bateman Manuscript Project, (McGraw-Hill, New York, 1953).
14. J.W. Goodman, "Introduction to Fourier Optics", (McGraw-Hill, New York, 1968).
15. M. Born and E. Wolf, "Principles of Optics", (Pergamon, New York, 1959).
16. J.H. Erkkila and M.E. Rogers, "Diffracted Fields in the Focal Volume of a Converging Wave," J. Opt. Soc. Am. 71, 904 (1981).
17. Y. Li and E. Wolf, "Focal Shifts in Diffracted Converging Spherical Waves," Opt. Comm. 39, 211 (1981).
18. A.L. Mikaelian, "Self-Focusing Media with Variable Index of Refraction," Prog. Opt. 17, 279 (1980).

CHAPTER 6

DIFFRACTION AND SELF-FOCUSING IN GRADED-INDEX MEDIA II:

SLOW AXIAL TAPERING

6.1 INTRODUCTION

The previous chapter on scalar diffraction in GRIN media gave a detailed study of the three-dimensional light distribution at the foci of a quadratic-index medium. The medium was assumed graded in the transverse direction only, and monochromatic plane and spherical waves were taken to be incident on a circular aperture at the endface of the medium. Remarkably simple expressions were obtained for the focused field, and the intensity half-widths at the foci. In the present chapter we generalize these results to the case of the medium having a slow axial gradient (tapering) in addition to the transverse gradient, by extending the modal technique outlined in section 5.2. This is formally carried out by using the adiabatic local modes to describe propagation within the medium, and the procedure is valid for any taper shape, provided it satisfies a slowness criterion. The correct generalizations of the formulae of chapter 5 are thus obtained.

An alternative approach to focusing and imaging in tapered media has been developed by Gomez-Reino et al.¹ and is based on a Green's function technique. Their approach does not, by necessity, assume slow tapering and consequently leads to a considerably more complex expression for the

diffracted field than is given here. Analytic results can only be obtained for a relatively small number of specific taper shapes,²⁻⁸ and information on the three-dimensional structure of the focused field is not readily obtained. The slow tapering limit of the Gomez-Reino theory gives results which are consistent with the method used here (see section 6.7) but lacks the conceptual simplicity of our approach.

Following this introduction, we give a brief description of the taper model in section 6.2. In section 6.3 we outline the adiabatic mode formalism, which is the basis of the scalar diffraction theory presented in section 6.4. We apply this diffraction formalism in section 6.5 to the analysis of the structure of the self-focused field for monochromatic plane wave incidence. Image transmission in the slowly tapered GRIN medium is briefly discussed in section 6.6. Section 6.7 compares some results with those obtained from the Gomez-Reino approach, and concluding remarks form section 6.8.

6.2 THE TAPERED QUADRATIC MEDIUM

The last chapter was concerned with the structure of the focused field in a quadratic medium:

$$n^2 = n_0^2 \{ 1 - r^2 / \rho_0^2 \} \quad (1)$$

where $r-\theta-z$ are cylindrical coordinates with z taken as the optical axis (see Fig. 1 of chapter 5), and n_0 and ρ_0 are constants.

In this chapter we show how the focused field is modified when the quadratic medium is slowly tapered:

$$n^2 = n_0^2 \{1 - r^2 / \rho^2(z)\} \quad (2)$$

Here $\rho(z)$, which we call the taper shape, is an arbitrary function of z (provided it satisfies a slowness criterion given later). For example, choosing

$$\rho(z) = \rho_0 (1 + z/D)^2 \quad (3)$$

(with ρ_0 , D constants) gives the parabolically tapered quadratic-index medium of chapters 2 and 3.

The diffraction problem is illustrated in Fig. 1 of chapter 5. The medium is graded (with index given by Eqn. (2)) in the $z > 0$ half-space, and an opaque screen with an aperture lies on the $z = 0$ plane. For convenience we take the medium in the $z < 0$ half-space to be homogeneous ($n^2 = n_1^2$). We investigate the diffracted field in the graded medium due to fields incident on the aperture from sources located in the $z < 0$ region. We assume an implicit time-dependence $\exp(-i\omega t)$, and that the optical field can be approximated by a scalar field satisfying the Helmholtz equation.

As discussed in section 5.2 of the last chapter, the natural length scale to use is the wavelength of light in the uniform medium of index n_0 , divided by 2π (i.e. $\lambda/2\pi n_0$ where λ denotes the free-space wavelength). Here, as in the last chapter, all length quantities (for example r , z and all others introduced later) are assumed to have been made dimensionless by scaling in this way.

With this length scaling, the Helmholtz equation becomes

$$\{\nabla^2 + (n^2 / n_o^2)\} E = 0 \quad (4)$$

6.3 ADIABATIC MODE FORMALISM

We proceed to describe the adiabatic local mode formalism and its limitations when applied to wave propagation in slowly tapered GRIN media.

The local modes are obtained from the well-known Laguerre-Gaussian modes⁹ of the untapered quadratic-index medium, described by Eqn. (1):

$$\psi_{\text{untap},jm} = r^j \exp[-r^2/2\rho_o] L_m^j[r^2/\rho_o] \frac{\cos}{\sin}(j\theta) \exp[i\beta_{jm}z] \\ j,m = 0,1,2,\dots \quad (5)$$

where β_{jm} , the propagation constants, are given by

$$\beta_{jm} = \sqrt{1 - 2(j+2m+1)/\rho_o} \quad (6)$$

(the reader is reminded of the length scaling implicit in Eqns. (5)-(6)).

Note that there are also backward propagating modes with $\exp[-i\beta_{jm}z]$ replacing the complex exponential in Eqn. (5).

From Eqns. (5)-(6), the local modes of the tapered quadratic medium, Eqn. (2), are obtained:⁹

$$\psi_{\text{local},jm} = r^j \exp[-r^2/2\rho(z)] L_m^j[r^2/\rho(z)] \frac{\cos}{\sin}(j\theta) \exp[i\phi_{jm}(z)] \quad (7)$$

with phase

$$\Phi_{jm}(z) = \int_0^z \beta_{jm}(s) ds \quad (8)$$

The local modes are clearly not exact solutions of the Helmholtz equation. In fact the exact solution would be a linear combination of all the local modes with z -dependent modal amplitudes. This is the approach of coupled-mode theory which describes the interchange of power between the various local modes at locations along the optical (z) axis. However, if the tapering is slow enough then a negligible amount of power will be exchanged by the local modes, and coupling between them can be neglected⁹. The local modes are then normalized to ensure that each propagates constant power (i.e. is adiabatic) and can then be treated as the exact modes of the medium.

Here we will take the modulus-squared of the scalar field to represent the time-averaged intensity of the optical field. Normalization of the local modes follows from the requirement that the total integrated intensity over any plane $z = z_a$ be independent of z_a . Hence the adiabatic local modes are given by:

$$\begin{aligned} \Psi_{jm} = & \left\{ 1 / \sqrt{\rho(z)} \right\}^{j+1} r^j \exp[-r^2/2\rho(z)] L_m^j [r^2/\rho(z)] \\ & \times \frac{\cos(j\theta)}{\sin(j\theta)} \exp[i\Phi_{jm}(z)] \end{aligned} \quad (9)$$

We now comment on the validity of individual adiabatic local modes in describing wave propagation in the tapered GRIN medium.

Firstly note that the adiabatic local modes have planar wavefronts (surfaces of constant phase). However, the actual field solutions inside tapered media must have curved wavefronts - for example, the parabolically tapered quadratic-index medium (chapter 3) supports guided wave propagation with spherical wavefronts centred on the apex of the taper. Only near the axis of the taper can the curved wavefronts always be adequately approximated by a plane. Thus an individual adiabatic mode can only be an accurate representation of the field if it is paraxial (i.e. field concentrated about the taper axis). The condition for this is easily obtained from the local propagation constants (see section 5.2):

$$2(j + 2m + 1)/\rho(z) \ll 1 \quad (10)$$

so that

$$\begin{aligned} \beta_{jm}(z) &= \sqrt{1 - 2(j + 2m + 1)/\rho(z)} \\ &\approx 1 - (j + 2m + 1)/\rho(z) \end{aligned} \quad (11)$$

Under the assumption of Eqn. (10) the phase of the adiabatic modes has the particularly simple form

$$\Phi_{jm}(z) = z - (j + 2m + 1) \gamma(z) \quad (12)$$

where

$$\gamma(z) = \int_0^z ds / \rho(s) \quad (13)$$

We require more than Eqn. (10) to ensure the validity of the adiabatic modes. Clearly the tapering must be slow enough to preclude significant power coupling between the local modes. This will be the case provided (see Snyder and Love⁹):

$$|z_B dp/dz| \ll \rho(z) \quad (14)$$

where z_B is the beat length between consecutive local modes. From Eqn. (12), $z_B = \pi z/\gamma(z)$ and hence the slowness criterion becomes

$$|[\pi z/\gamma(z)] dp/dz| \ll \rho(z) \quad (15)$$

The application of the two conditions, Eqns. (10) and (15), is demonstrated in appendix A for a parabolically tapered quadratic-index medium. Wave propagation inside this taper was analysed in chapter 3, and the validity of the adiabatic formalism can be checked directly.

6.4 SCALAR DIFFRACTION IN A SLOWLY TAPERED GRIN MEDIUM

The diffraction problem has been outlined in section 6.2. To simplify the discussion, we consider an axisymmetric system - a circular aperture of radius a (centred on the optical axis) and a θ -independent incident field. Following the usual approximation of scalar diffraction theory, the field on the aperture plane is written

$$E(r, z=0) = \begin{cases} \phi(r) & , r < a \\ 0 & , r > a \end{cases} \quad (16)$$

(i.e. we assume that the effect of the opaque screen is merely to truncate the incident field).

The diffracted field inside the graded medium is represented as a superposition of the adiabatic modes. By symmetry, only those forward propagating modes with $j = 0$ will be excited. Thus

$$E(r, z) = \sum_{m=0}^{\infty} a_m \Psi_{0m}(r, z) \quad (17)$$

Orthogonality of the adiabatic modes gives

$$a_m = \left(2/\sqrt{\rho_0}\right) \int_0^a s \phi(s) \exp(-s^2/2\rho_0) L_m(s^2/\rho_0) ds \quad (18)$$

The requirement that only those adiabatic modes satisfying Eqn. (10) be excited with significant power, places certain restrictions on the parameters of the optical system (see Appendix B, section 6.9.2).

Substitution of Eqn. (18) into Eqn. (17) and evaluation of the summation¹⁰ gives the diffracted field:

$$E(r, z) = \left[1 / i\bar{\sigma}(z)\zeta(z)\right] \exp\{i[z + r^2/2\bar{\tau}(z)]\} \times \int_0^{a\zeta(z)} s \phi[s/\zeta(z)] \exp[is^2/2\bar{\tau}(z)] J_0[rs/\bar{\sigma}(z)] ds \quad (19)$$

where

$$\bar{\sigma}(z) = \rho(z) \sin \gamma(z) \quad (20a)$$

$$\bar{\tau}(z) = \rho(z) \tan \gamma(z) \quad (20b)$$

$$\zeta(z) = \sqrt{\{\rho(z) / \rho_0\}} \quad (20c)$$

Equation (19) above is the generalization of Eqn. (5.20) of the last chapter (which describes the diffracted field in a transversely graded medium) to include the effects of slowly tapering the characteristic radius. Basic tapering effects originate from the integrated phase of the local modes, as defined in Eqn. (8).

6.5 SELF-FOCUSING OF PLANE WAVES

We now show how the self-focused field is modified by the slow axial tapering. We specifically consider the case of plane waves normally incident on the aperture, with unit intensity. Then

$$\phi [s / \zeta(z)] \equiv 1 \quad (21)$$

in Eqn. (19). We can now express Eqn. (19) explicitly in terms of the Lommel functions U_1 and U_2 (see Born and Wolf¹¹, or section 5.4) leading to the intensity distribution behind the aperture:

$$I(r,z) = \{1/\zeta(z)\cos\gamma(z)\}^2 \{U_1^2(\bar{p}, \bar{q}) + U_2^2(\bar{p}, \bar{q})\} \quad (22a)$$

with $\bar{p}$ and $\bar{q}$ given by

$$\bar{p} = -a^2\zeta^2(z)/\bar{\tau}(z) \quad (22b)$$

$$\bar{q} = a\zeta(z)/\bar{\sigma}(z) \quad (22c)$$

In the subsections that follow we consider some specific details of the focused field. Note that Appendix B (section 6.9.2) summarizes the approximations leading to the derivation of the scalar diffracted field.

6.5.1 Intensity Along the Optical Axis

The intensity distribution along the z -axis is most readily obtained from Eqns. (19)-(21):

$$I(r=0,z) = \{a^2 \zeta(z)/2\bar{\sigma}(z)\}^2 \operatorname{sinc}^2\{a^2 \zeta^2(z)/4\bar{\tau}(z)\} \quad (23)$$

Focusing occurs at locations given by the maxima of Eqn. (23). Note that the sinc^2 part of Eqn. (23) varies considerably more rapidly with z at optical frequencies, than the function multiplying it (recall the length scaling introduced in section 6.2 and implicit in Eqn. (23)). Thus the maxima of Eqn. (23) are approximately located at the values of z (say $z = \tilde{z}_m$) maximizing the sinc^2 function (and are exactly located there in the geometrical optics limit). We find that $\tilde{z}_m$ is given by the solutions to

$$\gamma(\tilde{z}_m) = (2m + 1) \pi/2 \quad (24)$$

(Here m is an integer labelling the different solutions). Thus the focal points are approximately located at $\tilde{z}_m$ and exactly located there in the geometrical optics limit.

Equation (24) could have been obtained more directly from ray theory. In a slowly tapering medium a quantity known as the adiabatic invariant¹²⁻¹³ is conserved, and leads to the following ray paths:

$$r = r_i \zeta(z) \cos \int_0^z ds / \sqrt{\rho^2(s) - r_i^2 \zeta^2(s)} \quad (25)$$

Equation (25) describes a meridional ray launched normal to the aperture, a distance r_i from its centre (i.e. plane waves normally incident). If only paraxial rays are launched (satisfying $r_i^2 \zeta^2(z) \ll \rho^2(z)$) then all the rays will cross the optical axis at the locations $\tilde{z}_m$ in Eqn. (24) - the geometrical optics foci.

The width of the focused field about the foci can be obtained by making the following change of variable

$$z = \tilde{z}_m + z' \quad (26)$$

in Eqn. (23). Expanding the argument of the sinc^2 function to first order in z' and neglecting the z' dependence of the function multiplying the sinc^2 , we obtain

$$I \approx \{a^2 \zeta(\tilde{z}_m)/2\rho(\tilde{z}_m)\}^2 \text{sinc}^2\{a^2 z'/4\rho_o \rho(\tilde{z}_m)\} \quad (27)$$

Following chapter 5, we define the intensity half-width, Ω_T , as the value of z' for which the first intensity zero is obtained. From Eqn. (27),

$$\Omega_T = 4\pi \rho_o \rho(\tilde{z}_m)/a^2 \quad (28)$$

Letting Ω_P denote the intensity half-width at the foci of an untapered quadratic medium (see section 5.4.1), then

$$\Omega_T / \Omega_P = \zeta^2(\tilde{z}_m) \quad (29a)$$

where ζ is defined in Eqn. (20c). The above equation may be rewritten

$$\Omega_T/\rho(\tilde{z}_m) = \Omega_P/\rho_0 \quad (29b)$$

to emphasize the simple scaling of the intensity width that results from the tapering.

6.5.2 Intensity in Geometrical Focal Plane

The intensity distribution in the geometrical focal plane $z = \tilde{z}_m$ is exactly a scaled version of the Airy intensity pattern found in an untapered quadratic medium (see section 5.4.2):

$$I(r, z=\tilde{z}_m) = \{a^2 \zeta(\tilde{z}_m)/2\rho(\tilde{z}_m)\}^2 \{2J_1(u)/u\}^2 \quad (30a)$$

with u given by

$$u = \zeta(\tilde{z}_m) ar/\rho(\tilde{z}_m) \quad (30b)$$

(This is in contrast to the intensity distribution along the optical axis, which is clearly not a simple scaled version of the untapered distribution.)

Defining the half-width of the Airy pattern, ω_T , to be the radius of the first intensity zero, then

$$\omega_T = 1.22\pi \rho(\tilde{z}_m) / a \zeta(\tilde{z}_m) \quad (31)$$

With ω_P denoting the half-width of the Airy pattern in an untapered quadratic medium (see Eqn. (5.39)) then

$$\omega_T / \omega_P = \zeta(\tilde{z}_m) \quad (32)$$

Equations (29) and (32) reveal that the effect of slow tapering is to spread out (or concentrate - depending on the nature of the tapering) the focused intensity, by the factor ζ^2 along the optical axis, and by the factor ζ orthogonal to it, where ζ^2 is the radius ratio defined in Eqn. (20c). The intensity at the geometrical focal point changes by the factor $1/\zeta$ (see Eqn. (23) or (30)).

6.6 IMAGE TRANSMISSION

Gomez-Reino et al.¹⁻⁸ have shown that tapered GRIN media can transmit paraxial images, and have obtained analytic solutions for several taper shapes. Here we comment on imaging inside a slowly tapered GRIN medium, using the adiabatic mode formalism.

Suppose there is an arbitrary field distribution (the object field distribution) on the $z = 0$ plane (see Fig. 1 of chapter 5). Representing the field as a superposition of adiabatic modes (Eqn. (9)) we find that image planes are located at $\tilde{z}_{\text{image},m}$ - the solutions of:

$$\gamma(\tilde{z}_{\text{image},m}) = m\pi \quad (33)$$

and the magnification factor is

$$M_m = \zeta(\tilde{z}_{\text{image},m}) \quad (34)$$

In Eqns. (33)-(34), m is an integer used to label the image planes. The even values of m correspond to planes on which the propagating adiabatic modes re-establish the same relative phases that they began with (at $z = 0$), leading to an erect image. Odd values of m correspond to the

images inverted through the point of intersection of the optical axis with the image plane.

Comparisons with the Gomez-Reino approach are discussed in the next section.

6.7 COMPARISON WITH NON-ADIABATIC THEORY

The approach developed by Gomez-Reino et al.¹ does not by necessity assume slowness of tapering, and we shall refer to it as the non-adiabatic theory.

Comparison of adiabatic and non-adiabatic approaches is demonstrated in tables I-III for three different taper shapes: a linear taper $\rho(z) = \rho_0(1 + z/D)$ (table I), a parabolic taper $\rho(z) = \rho_0(1 + z/D)^2$ (table II), and an exponential taper $\rho(z) = \rho_0 \exp(z/D)$ (table III). In each case ρ_0 and D are constants, and

$$Q \equiv D / \rho_0 \quad (35)$$

determines the degree of tapering. Comparisons are made between locations of focal and image planes and image magnifications. The slowness criterion (Eqn.(15)) provides the link between the two sets of results, and ensures that the adiabatic and non-adiabatic approaches are in agreement.

For example, in the case of the linear taper (table I), a necessary and sufficient condition for the slowness criterion to be satisfied everywhere is $Q \gg \pi$. Under this condition, the two sets of results in table I are clearly equivalent.

For the parabolic taper (table II) the slowness criterion is $Q \gg 2\pi$, and once again, the two sets of results in table II are equivalent.

I. $\rho(z) = \rho_0 (1 + z/D)$		
	Gomez-Reino et al. ³	Adiabatic Theory
Focal Planes	$\sqrt{\{Q^2 - (1/4)\}} \log(1 + \tilde{z}_m/D)$ $= \cot^{-1} \{1/2\sqrt{\{Q^2 - (1/4)\}}\}$	$Q \log(1 + \tilde{z}_m/D)$ $= (2m + 1) \pi/2$
Image Planes	$(1 + \tilde{z}_{\text{image},m}/D)$ $= \exp\{m\pi/\sqrt{\{Q^2 - (1/4)\}}\}$	$(1 + \tilde{z}_{\text{image},m}/D)$ $= \exp\{m\pi/Q\}$
Image Magnification	$M_m = \sqrt{\{1 + \tilde{z}_{\text{image},m}/D\}}$ $= \exp\{m\pi/2\sqrt{\{Q^2 - (1/4)\}}\}$	$M_m = \sqrt{\{1 + \tilde{z}_{\text{image},m}/D\}}$ $= \exp\{m\pi/2Q\}$

Table I. Comparison of adiabatic and non-adiabatic theories for linear taper.

II. $\rho(z) = \rho_0 (1 + z/D)^2$		
	Gomez-Reino et al. ⁵	Adiabatic Theory
Focal Planes	$Q \tilde{z}_m / (D + \tilde{z}_m)$ $= \cot^{-1} (1/Q)$	$Q \tilde{z}_m / (D + \tilde{z}_m)$ $= (2m + 1) \pi/2$
Image Planes	$Q \tilde{z}_{\text{image},m} / (D + \tilde{z}_{\text{image},m})$ $= m\pi$	$Q \tilde{z}_{\text{image},m} / (D + \tilde{z}_{\text{image},m})$ $= m\pi$
Image Magnification	$M_m = 1 + \tilde{z}_{\text{image},m}/D$	$M_m = 1 + \tilde{z}_{\text{image},m}/D$

Table II. Comparison of adiabatic and non-adiabatic theories for quadratic taper.

III. $\rho(z) = \rho_0 \exp(z/D)$		
	Gomez-Reino et al. ⁷	Adiabatic Theory
Focal Planes	$Y_1(Q) J_0(Q \exp[-\tilde{z}_m/D])$ $= J_1(Q) Y_0(Q \exp[-\tilde{z}_m/D])$	$\exp[-\tilde{z}_m/D]$ $= 1 - (1/Q)(2m + 1) \pi/2$
Image Planes	$Y_0(Q) J_0(Q \exp[-\tilde{z}_{\text{image},m}/D])$ $= J_0(Q) Y_0(Q \exp[-\tilde{z}_{\text{image},m}/D])$	$\exp[-\tilde{z}_{\text{image},m}/D]$ $= 1 - m\pi/Q$
Image Magnification	$M_m = \frac{Y_0(Q \exp[-\tilde{z}_{\text{image},m}/D])}{Y_0(Q)}$	$M_m = \frac{1}{\sqrt{1 - m\pi/Q}}$

Table III. Comparison of adiabatic and non-adiabatic theories for exponential taper

In the case of the exponential taper (table III), a necessary and sufficient condition for the slowness criterion to be satisfied is -

$$Q \exp[-z/D] \gg \pi \quad (36)$$

Unlike the linear and parabolic tapers, the exponential taper cannot satisfy the slowness criterion everywhere - it always breaks down for large enough z . However, provided $Q \gg \pi$ then Eqn. (36) will be satisfied for small enough values of z , and we can replace the Bessel functions in table III with their large argument asymptotic formulae:¹⁴

$$J_\nu(x) \approx \sqrt{2/\pi x} \cos\{x - \nu\pi/2 - \pi/4\} \quad (37)$$

$$(|x| \rightarrow \infty)$$

and similarly for $Y_\nu(x)$. Agreement between the two sets of results in table III then follows.

6.8 CONCLUSION

We have generalized the results of the last chapter, concerning the structure of the focused field in a transversely graded medium, to the case in which the medium is slowly, but otherwise arbitrarily, tapered. Our approach, an extension of the method used in chapter 5, is based on an adiabatic local mode representation of the field, and leads to a remarkably simple expression for the focused field. As a result of the tapering, the intensity near focus is spread out (or concentrated) by the factor ζ^2 along the optical axis, and ζ in an orthogonal direction, where ζ^2 is the radius ratio of Eqn. (20c). Although the scaling of the intensity around the focus depends only on the local radius, the actual position of

the foci (and the image planes) depends on the integrated effects of tapering the medium.

We emphasize the conceptual simplicity of this approach, which comes about by assuming slowness at the onset. An alternative approach is to obtain approximate solutions (in the limit of slow tapering) to the non-adiabatic Gomez-Reino theory,¹⁵ and this has been applied to study the self-focused field in tapered GRIN media.¹⁶ Results obtained through this Greens function treatment are in complete agreement with those given here.

6.9.1 APPENDIX A: ADIABATIC MODES IN A PARABOLICALLY TAPERED QUADRATIC MEDIUM

Here we give a specific example of adiabatic mode propagation in tapered media. Wave propagation in the parabolically tapered quadratic index medium has been described in some detail in chapter 3. To a very good approximation, the paraxial fields (referred to as basis fields) can be written in the form

$$\psi_{jm}^{\text{basis}} = \{1/\sqrt{\rho(z)}\}^{j+1/2} r^j \exp[-r^2/2\rho(z)] L_m^j[r^2/\rho(z)] \quad (\text{A1})$$

$$\times \frac{\cos(j\theta)}{\sin(j\theta)} \mathcal{H}_{\mu}^{(1),(2)}(D+z)$$

where μ is given by

$$\mu = (1/2) \sqrt{1 + 8(j+2m+1) D^2/\rho_0} \quad (\text{A2})$$

In Eqn. (A1), $\mathcal{H}_{\mu}^{(1)}$ and $\mathcal{H}_{\mu}^{(2)}$ are Hankel functions¹⁴ describing forward and backward propagating waves respectively, and the length scaling introduced in section 6.2 has been applied here.

By comparing Eqns. (9) and (A1) we see that the adiabatic local modes differ from the basis fields only in their propagation characteristics:

$$\begin{aligned} H_{\mu}^{(1),(2)}(D+z), & \quad \text{for basis fields} \\ e^{\pm i\phi_{jm}(z)} / (D+z)^{1/2}, & \quad \text{for adiabatic modes} \end{aligned} \quad (A3)$$

(Note that in Eqn. (A3) we have included backward propagating solutions.)
From Eqn. (8) we obtain

$$\phi_{jm}(z) = C_1 + \sqrt{(D+z)^2 - v^2} - v \sec^{-1}\{(D+z)/v\} \quad (A4)$$

where C_1 is a constant independent of z , and $v = \sqrt{\mu^2 - (1/4)}$.

It is now a simple matter to apply the conditions in Eqns. (10) and (15) to the basis fields, to show that they are well represented by the adiabatic modes. The slowness criterion (Eqn. (15)) is $D / \rho_0 \gg 2\pi$ and we can also assume $D > 1$ (recall the implicit scaling of all length quantities). Hence (see Eqn. (A2))

$$\mu \gg 1 \quad (A5)$$

and from the defining equation for v (see above) we have $v \approx \mu$. The paraxial condition (Eqn. (10)) is $(D+z) \gg v$ and thus

$$(D+z) \gg \mu \quad (A6)$$

At this point we introduce a well-known asymptotic formula for Hankel functions¹⁴ -

$$\mathcal{H}_{\mu}^{(1),(2)}(\mu \sec \delta) \approx \sqrt{2/\pi \mu \tan \delta} \exp\{\pm i[\mu \tan \delta - \mu \delta - \pi/4]\} \quad (\mu \gg 1) \quad (\text{A7})$$

Equations (A5) and (A6) allow us to apply this asymptotic formula to write

$$\begin{aligned} \sqrt{(\pi/2)(D+z)} \mathcal{H}_{\mu}^{(1),(2)}(D+z) \\ \approx \exp\{\pm i[\sqrt{(D+z)^2 - v^2} - v \sec^{-1}[(D+z)/v] - \pi/4]\} \end{aligned} \quad (\text{A8})$$

(we have also made use of the relation $\mu \approx v$ in the RHS of Eqn. (A8)). Comparison of Eqns. (A8), (A3) and (A4) clearly reveals that the j m-th adiabatic mode is an accurate representation of the j m-th basis field.

6.9.2. APPENDIX B: VALIDITY OF THE DIFFRACTION ANALYSIS

In this Appendix we summarize the approximations used to derive the scalar diffracted field, Eqns. (19)-(20), to give some appreciation for its range of validity.

Using the adiabatic local modes to describe propagation within the tapered medium is valid provided the slowness criterion (Eqn. (15)) is satisfied, and provided only those modes satisfying Eqn. (10) are excited. The latter condition is introduced into the modal representation of the scalar field (Eqn. (17)) by requiring that the spectrum of modal amplitudes, excited by the source, decrease to zero very rapidly with increasing mode number, so that only those modes satisfying Eqn. (10) have significant power. This results in the imposition of restrictions on the parameters of the optical system. A simple example is given below.

In the case of plane waves normally incident on the aperture, and a taper shape satisfying $\rho(z) \geq \rho(0) = \rho_0$, then the same argument as used in section 5.5.1 leads to the following restrictions on the parameters a and ρ_0 :

$$\frac{\tanh(2/\rho_0)}{(2/\rho_0)} \ll a^2/2 \ll \min \left\{ \begin{array}{l} [\coth(2/\rho_0)] / (2/\rho_0) \\ 2 / (2/\rho_0)^2 \end{array} \right\} \quad (B1)$$

Equation (B1) also incorporates the condition $a^2/\rho_0^2 \ll 1$ which is necessary for the quadratic-index model (Eqn. (2)) to be physically realistic.

As well as restrictions on the parameters, there are also restrictions on the magnitude of z . This results from the neglect of higher order terms in the expansion of the local propagation constant (Eqn. (11)) when used to determine the phase of the adiabatic mode (Eqn. (8)). This will not lead to significant error provided -

$$\left\{ (2m + 1)^2/2 \right\} \int_0^z ds / \rho^2(s) \ll \pi \quad (B2)$$

In the simple case of normally incident plane waves and a taper shape satisfying $\rho(z) \geq \rho(0) = \rho_0$, then it can be shown using the arguments of section 5.5.1 that the order of magnitude of the largest mode number excited by the incident field is

$$\Delta A \sim \begin{cases} 1 / [2 \tanh^{-1}(a^2/\rho_0)] , & a^2/\rho_0 < 1 \\ 1 / [2 \coth^{-1}(a^2/\rho_0)] , & a^2/\rho_0 > 1 \end{cases} \quad (B3)$$

The worst case in Eqn. (B2) occurs when $m \sim \Delta A$, leading to the restriction

$$\int_0^z ds/\rho^2(s) \ll 2\pi T^2(a^2/\rho_0) \quad (B4)$$

where the function T is defined by

$$T(x) = \begin{cases} \tanh^{-1}(x) / [1 + \tanh^{-1}(x)] , & x \leq 1 \\ \coth^{-1}(x) / [1 + \coth^{-1}(x)] , & x > 1 \end{cases} \quad (B5)$$

6.10 REFERENCES

1. C. Gomez-Reino and E. Larrea, "Imaging and Transforming Transmission through a Medium with Nonrotation-Symmetric Gradient-Index", *Appl. Opt.* 22, 387 (1983).
2. C. Gomez-Reino, E. Larrea, M.V. Perez and J.M. Cuadrado, "Transmittance Function and Modal Propagation in a Conical Gradient-Index Rod", *Appl. Opt.* 23, 1107 (1984).
3. C. Gomez-Reino, M.V. Perez, J. Sochacki and M. Sochacka, "Image and Transform Transmission through Divergent Conical GRIN Rods", *Opt. Commun.* 55, 5 (1985).
4. C. Gomez-Reino, M.V. Perez, J. Sochacki and M. Sochacka, "Image and Transform Transmission through Divergent Parabolical GRIN Rods", *Opt. Commun.* 55, 8 (1985).

5. C. Gomez-Reino and J. Sochacki, "Imaging and Transforming Capabilities of GRIN Rods with Noncylindrical Surfaces of Constant Index: a family of Exact Solutions", Appl. Opt. 24, 4375 (1985).
6. C. Gomez-Reino, E. Larrea, M.V. Perez and J.M. Cuadrado, "Imaging, Transforming and Modal Propagation in a Parabolic Gradient-Index Rod", Appl. Opt. 24, 4379 (1985).
7. C. Gomez-Reino, J. Linares and J. Sochacki, "Image and Transform Transmission through a GRIN Rod with Exponential Variation of Equi-Index Surfaces", Appl. Opt. 25, 1076 (1986).
8. C. Gomez-Reino, J. Linares and E. Larrea, "Imaging and Transforming Transmission through Tapered Gradient-Index Rods: Analytical Solutions", J. Opt. Soc. Am. A3, 1604 (1986).
9. A.W. Snyder and J.D. Love, "Optical Waveguide Theory", (Chapman and Hall, London, 1983).
10. A. Erdelyi, Ed., "Higher Transcendental Functions", Vol. 2, Bateman Manuscript Project (McGraw-Hill, New York, 1953).
11. M. Born and E. Wolf, "Principles of Optics", (Pergamon, New York, 1959).
12. A. Ankiewicz, C. Pask and A.W. Snyder, "Slowly Varying Optical Fibers", J. Opt. Soc. Am. 72, 198 (1982).

13. D.M. Milder, "Ray and Wave Invariants for SOFAR Channel Propagation", J.Acoust. Soc. Am. 46, 1259 (1969).
14. M. Abramowitz and I.A.Stegun, Eds., "Handbook of Mathematical Functions", (Dover, New York, 1972).
15. C. Gomez-Reino, M.V. Perez and E. Larrea, "Imaging and Transforming Transmission through Inhomogeneous Media with Revolution Symmetry", Opt. Commun. 44, 8 (1982).
16. C. Gomez-Reino and E. Larrea, "Pupil Effect in GRIN material", Appl. Opt. 22, 970 (1983).

CHAPTER 7

AN APPLICATION IN OPTICAL ELECTRONICS:

SPLICING OF OPTICAL WAVEGUIDES WITH LOSSLESS GRIN TAPERS

7.1 INTRODUCTION

Chapter 4 gave an exact ray analysis of a strictly adiabatic taper, and the light compression properties of the taper were examined using concentrator theory. Here we discuss an application which is closely related to the concentration problem - how to couple light from a large diameter optical waveguide into a smaller diameter waveguide, in such a way as to minimize loss. In optoelectronic systems there often arises the need to join together two single-mode fibers of different radii, so that radiation losses are minimized. One scheme is to slowly taper one of the fibers to match the cross-section of the other. If the tapering is adiabatically slow then radiation loss along the taper will be small¹, but an impractically long length of taper is often required. Another approach is to use a cylindrical GRIN rod to expand the (approximately) Gaussian field distribution of the fundamental mode of the fiber².

In this chapter, we investigate the application of strictly adiabatic tapers as waveguide connectors (originally suggested by Marcatili³). Losses occur only at the waveguide/taper junction, and are due, principally, to the mismatch between the curved wavefronts of the taper fields and the planar wavefronts of the waveguide modes. We give a simple expression for the coupling efficiency, for perfect alignment of

taper and waveguides. For convenience we consider the specific case of the simplest strictly adiabatic taper - a linearly tapered slab with radial grading - joining together dielectric slab waveguides of different widths.

In section 7.2 we compare the modes of the 2D taper with the modes of a dielectric slab waveguide. Basic similarities suggest the taper as a mode expander (or concentrator) for the slab, and the different wavefront curvatures is the primary source of loss. Further discussion is restricted to the simplest case: single-moded waveguides and taper, and coupling of fundamental modes. In section 7.3 we derive a Gaussian approximation for the fundamental mode of the taper, and use this in section 7.4 to obtain a simple formula for the coupling efficiency. This formula shows clearly the contribution to the loss due to mismatch of the spot-sizes and wavefront curvatures at the waveguide/taper junctions, and is analysed in detail in section 7.5. Concluding remarks form section 7.6.

Throughout this chapter we use scalar fields; under the weak-guidance approximation (see section 1.5) the full electromagnetic fields can be constructed from solutions of the scalar wave equation. Furthermore, we assume an implicit time dependence $\exp(-i\omega t)$, where $\omega = ck = 2\pi c/\lambda$ and λ is the free-space wavelength.

7.2 THE MARCATILI TAPER AND THE DIELECTRIC SLAB

The 2D taper model was described in chapter 4. Recall that the refractive index distribution is given by:

$$n^2 = \begin{cases} n_o^2 [1 + N_o^2/R^2] , & |\theta| < \theta_o \\ n_o^2 , & \theta_o < |\theta| < \pi \end{cases} \quad (1)$$

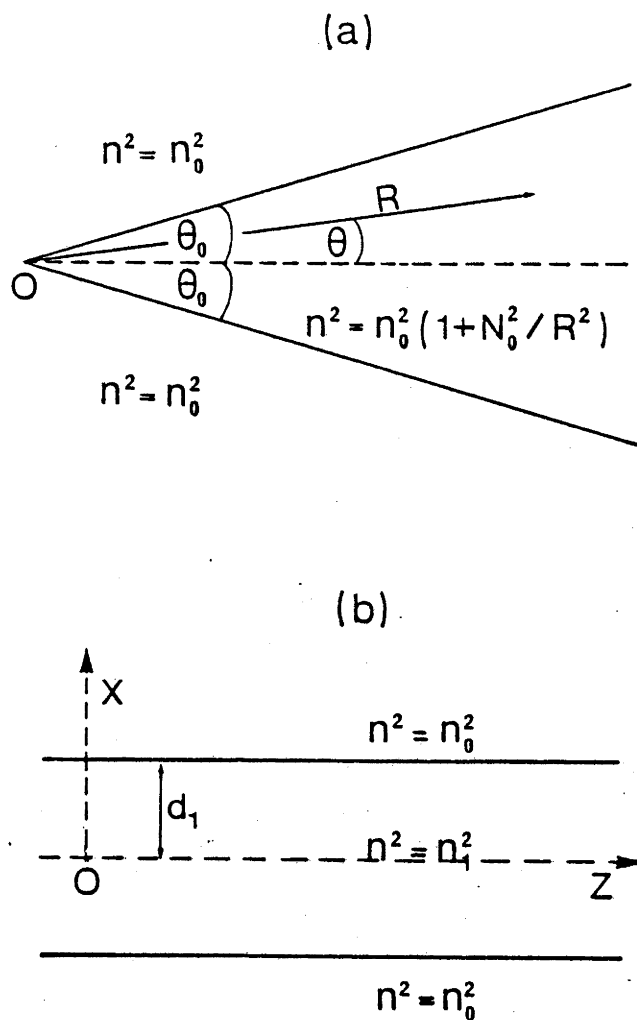


Fig. 1 Co-ordinate systems and refractive index distributions for (a) Marcatili taper, and (b) dielectric slab waveguide.

where $R - \theta$ are plane-polar coordinates (see Fig. 1(a)), N_0 is a constant (with dimensions of length), and n_0 is the uniform refractive index of the medium outside the taper (the cladding). As emphasized by Marcatili³, the taper is characterized by a constant V (waveguide parameter) if we re-define this quantity over a radial arc (see later this section).

The bound modes of the taper are closely related to the bound modes of a step-index slab waveguide. A dielectric slab of refractive index n_1 and width $2 d_1$ surrounded by a uniform medium of index n_0 (see Fig. 1(b)) supports modes of the form

$$E = \psi(X) \exp(\pm i\beta Z) \quad (2)$$

where β is the propagation constant. For the even bound modes¹:

$$\psi(X) = \begin{cases} \cos[UX/d_1] & , \quad |X| \leq d_1 \\ \cos[U] \exp[-W(|X/d_1| - 1)] & , \quad d_1 \leq |X| \end{cases} \quad (3)$$

(and a similar expression for the odd modes). Here U and W are the modal parameters defined by

$$U = d_1(k^2 n_1^2 - \beta^2)^{1/2} \quad (4a)$$

$$W = d_1(\beta^2 - k^2 n_0^2)^{1/2} \quad (4b)$$

The waveguide is conveniently characterized by V , the waveguide parameter, defined by

$$\begin{aligned} V &= (U^2 + W^2)^{1/2} \\ &= k d_1(n_1^2 - n_0^2)^{1/2} \end{aligned} \quad (4c)$$

The eigenvalue equation for the (even) modes of Eqn (3) is

$$U \tan U = W \quad (5)$$

Odd and even modes are labelled by the mode number, N , and the cut-off for mode N occurs at

$$V = N\pi/2 \quad (6)$$

Note that the fundamental mode corresponds to $N = 0$, and is never cut-off.

By comparison, the bound modes of the taper are of the form (see Marcatili³, or section 4.6):

$$E = B(\theta) \mathcal{H}_{i\sqrt{v}}^{(1),(2)}(k n_o R) \quad (7)$$

where $\mathcal{H}_{i\sqrt{v}}^{(1)}(k n_o R)$ and $\mathcal{H}_{i\sqrt{v}}^{(2)}(k n_o R)$ are Hankel functions⁴ of pure imaginary order, describing propagation away from and towards the taper apex, respectively, and v is an eigenvalue. In Eqn (7), $B(\theta)$ gives the field distribution over a radial arc, and the wavefronts (surfaces of constant phase) are circles centred on the taper apex. It is important to note that the bound modes of the slab waveguide have planar wavefronts (see Eqn (2)).

It is convenient to introduce modal and waveguide parameters for the taper, analogous to Eqns (4) for the slab, but defined more naturally over a radial arc:

$$U_R = (R\theta_o) (k^2 n^2(R) - \beta_R^2)^{1/2} \quad (8a)$$

$$W_R = (R\theta_o) (\beta_R^2 - k^2 n_o^2)^{1/2} \quad (8b)$$

$$\begin{aligned}
V_R &= (U_R^2 + W_R^2)^{1/2} \\
&= k(R\theta_o) (n^2(R) - n_o^2)^{1/2} \\
&= k n_o N_o \theta_o \quad (8c)
\end{aligned}$$

Here $n(R)$ is the refractive index in the core of the taper, and β_R can be viewed as a local (radial) propagation constant, and is related to the order of the Hankel function in Eqn (7) through:

$$|v| = R^2(\beta_R^2 - k^2 n_o^2) \quad (9)$$

(Note that for a particular taper mode the RHS of Eqn (9) is an invariant.)

Equations (8) define natural parameters for the modes of the Marcatili taper. Substituting Eqns (1) and (9) into Eqns (8) gives

$$U_R = \theta_o (k^2 n_o^2 N_o^2 - |v|)^{1/2}$$

$$W_R = \theta_o \sqrt{|v|}$$

Thus, all three parameters, V_R , U_R and W_R , are independent of the radial distance from the taper apex to the arc over which they are defined.

The even bound modes of the taper may be written in terms of these parameters:

$$B(\theta) = \begin{cases} \cos[U_R(R\theta) / (R\theta_o)] & , \quad |\theta| < \theta_o \\ \cos[U_R] \frac{\cosh[W_R(\pi - |\theta|)/\theta_o]}{\cosh[W_R(\pi - \theta_o)/\theta_o]} & , \quad \theta_o < |\theta| < \pi \end{cases} \quad (10)$$

to emphasize the similarity with the bound modes of the slab (compare with Eqn. (3)). The eigenvalue equation for the above even modes is

$$U_R \tan U_R = W_R \tanh[W_R(\pi - \theta_o) / \theta_o] \quad (11)$$

It is easily shown that the odd/even bound mode with mode number N is cut-off at

$$V_R = N \pi / 2 \quad (12)$$

(compare with Eqn (6)). Note that bound modes have $kn_o < \beta_R < kn(R)$ and cut-off occurs when $\beta_R = kn_o$. (This follows from Eqns. (4.33) and (4.37) of chapter 4.)

The similarity between Eqns (3) and (10) is even more striking for taper modes with $W_R(\pi - \theta_o)/\theta_o \gg 1$ (i.e. modes tightly bound to the core of the taper). In this case the fields just outside the taper faces (at $\theta = \pm \theta_o$) may be approximated as follows:

$$\begin{aligned} \cos[U_R] & \frac{\cosh[W_R(\pi - |\theta|)/\theta_o]}{\cosh[W_R(\pi - \theta_o)/\theta_o]} \\ & \approx \cos[U_R] \exp[-W_R(|R\theta/R\theta_o| - 1)]. \end{aligned} \quad (13)$$

This close relationship between the field distributions on a surface of constant phase for the slab and taper is a consequence of both fields ($\psi(X)$ and $B(\theta)$) satisfying the same differential equation. Only the boundary conditions are different ($\psi(\pm\infty) = 0$ for the slab, and

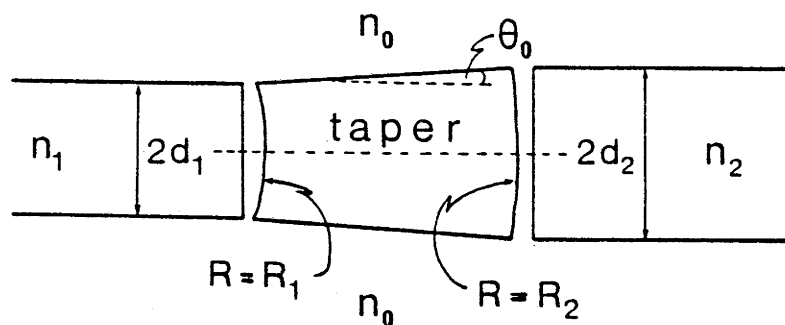


Fig. 2 Marcattili taper connecting two slab waveguides in a uniform medium of refractive index n_0 . The taper end-faces are the radial arcs $R = R_1$ and $R = R_2$.

$B(\pi) = B(-\pi)$, $dB/d\theta(\pi) = dB/d\theta(-\pi)$ for the taper) and this results in the slightly different cladding fields and eigenvalue equations. Note that in many situations of practical interest the modal fields will be tightly bound to the core regions and the cladding fields will be negligible.

These observations suggest the suitability of the Marcattili taper as a slab waveguide connector (shown schematically in Fig. 2). Apart from wavefront curvature, the field of any slab mode can be matched to a corresponding taper mode by appropriate design, and the taper can be regarded as a mode expander (or concentrator) for the slab waveguide. The principal losses will arise from the mismatch in wavefront curvature, and this will be discussed in the remainder of this chapter. To ensure that the basic physics is not obscured by mathematical complexities, we analyse the simple case of single-moded waveguides and taper, with fundamental modes approximated by Gaussian field distributions. We briefly develop the Gaussian approximation in the next section, before returning to the coupling problem.

7.3 GAUSSIAN APPROXIMATION FOR THE FUNDAMENTAL MODE

The idea behind the Gaussian approximation for the fundamental mode of the taper is a straightforward extension of the same concept used in fiber optics^{1,5}. The variational approach discussed in references 1 and 5, leads to the following stationary expression for the eigenvalue v :

$$v = \frac{\int_{-\pi}^{\pi} (dB/d\theta)^2 d\theta - k^2 n_o^2 N_o^2 \int_{-\theta_o}^{\theta_o} B^2(\theta) d\theta}{\int_{-\pi}^{\pi} B^2(\theta) d\theta} \quad (14)$$

(This is derived from the differential equation for $B(\theta)$, the general form of which is given in Eqn. (3.11).) We approximate the fundamental mode by a Gaussian function

$$B(\theta) \approx \exp[-\theta^2/2\theta_s^2] \quad (15)$$

where θ_s , the angular spot-size, is obtained by substituting Eqn (15) into Eqn (14) and setting $dv/d\theta_s = 0$. In any situation of practical interest, the taper angle will be small, and $\theta_s \ll \pi$. Thus, we can extend the limits of the θ -integrations, in Eqn (14), from $\pm \pi$ to $\pm \infty$ with negligible error. (For example, the fractional error to the B^2 integral is $\text{erfc}(\pi/\theta_s) \sim (\theta_s/\pi)^{3/2} \exp[-(\pi/\theta_s)^2]$ where erfc is the complementary error function.) This leads to the result:

$$V_R^2 = (\pi/4)^{1/2} (\theta_o/\theta_s) \exp(\theta_o/\theta_s)^2 \quad (16a)$$

Figure 3 shows the normalized spot-size, θ_s/θ_o , plotted against V_R . Note that the taper is single-moded when $V_R < \pi/2$ (see Eqn (12)). We write

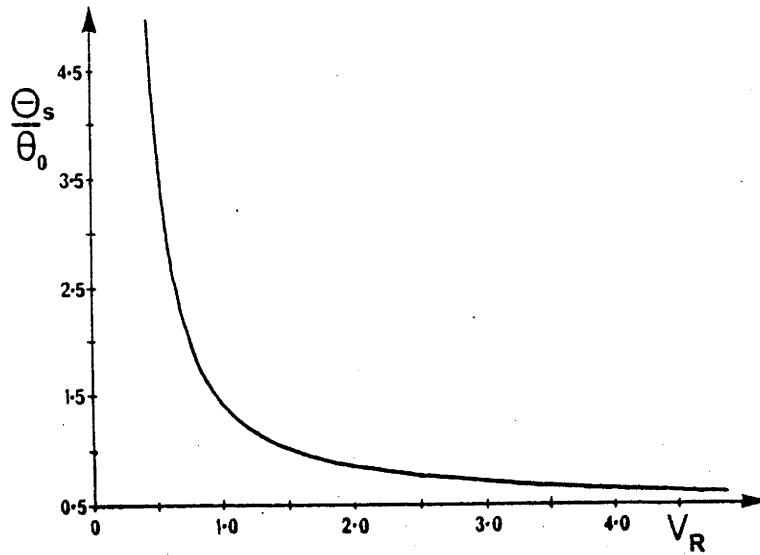


Fig. 3 Normalized spot-size, Θ_s/Θ_0 , versus waveguide parameter V_R for the fundamental taper mode. The same graph gives normalized spot-size for the slab (X_1/d_1) versus slab waveguide parameter (V).

$$\Theta_s = \Theta_0 f(V_R) \quad (16b)$$

where $f(V_R)$ is the function plotted in Fig. 3.

The Gaussian approximation for the fundamental mode of the slab waveguide (Fig. (1b)) is

$$\psi(X) \approx \exp[-X^2/2X_1^2] \quad (17)$$

where X_1 is the spot-size. The variational approach gives an identical spot-size equation:

$$V^2 = (\pi/4)^{1/2} (d_1/X_1) \exp(d_1/X_1)^2 \quad (18a)$$

or

$$X_1 = d_1 f(V) \quad (18b)$$

As pointed out in section 7.2, $\psi(X)$ and $B(\theta)$ satisfy the same differential equation but with different boundary conditions. However, the assumption that $\theta_s \ll \pi$ and extending the limits of integration means, physically, that boundary effects are negligible.

7.3.1 Taper design: Spot-size matching

We consider the problem of designing a taper to best connect the two waveguides of Fig. 2. Here slab #1 has index n_1 and width $2d_1$, and slab #2 has index n_2 and width $2d_2$. The cladding is uniform everywhere with refractive index n_0 . We assume that only fundamental modes are involved and they are all well approximated by Gaussian functions. If we neglect wavefront curvature, then we can simply match the spot-sizes at the two waveguide/taper junctions:

$$\begin{aligned} R_1 \theta_s &= X_1, \\ R_2 \theta_s &= X_2, \end{aligned} \tag{19a}$$

where X_1 and X_2 are the spot-sizes for slabs #1 and #2, respectively. Alternatively, if V_1 and V_2 are the waveguide parameters for the two slabs, then we can use Eqns. (16b) and (18b) to rewrite the above:

$$\begin{aligned} (R_1 \theta_o) f(V_R) &= d_1 f(V_1), \\ (R_2 \theta_o) f(V_R) &= d_2 f(V_2). \end{aligned} \tag{19b}$$

where f is the function plotted in Fig. 3. Any two slab waveguides can be matched using this approach, and only two restrictions are placed on the taper parameters (Eqns. (19a) or (19b)) thus giving some freedom in the design to minimize total losses. In the next section we derive a formula

for the device efficiency, which includes wavefront curvature, and allows us to choose all parameters to optimize the performance.

7.4 COUPLING EFFICIENCY FORMULA

Here we derive a simple formula for the coupling efficiency for the fundamental modes of two slabs joined by a Marcatili taper. This formula shows very clearly the loss due to mismatch of spot-sizes, and also due to the wavefront curvature.

The result can be compared with the coupling efficiency, η_B , for a perfectly aligned butt-joint of two slabs with fundamental mode spot-sizes x_1 and x_2 :

$$\begin{aligned}\eta_B &= 2 x_1 x_2 / (x_1^2 + x_2^2) \\ &= 2(x_1/x_2) / [1 + (x_1/x_2)^2]\end{aligned}\quad (20)$$

(see Marcuse⁶).

7.4.1 Preliminaries

The geometry is illustrated in Fig. 4. To facilitate the mode excitation calculation we use a curved interface (radius R_1) for the first slab/taper junction, and a planar interface (at $Z = Z_2$) for the second. (Note that the taper modes are orthogonal over a circle, the slab modes over a plane.)

We consider the simple case (neglecting backscatter) of the fundamental mode of slab #1 coupling into the fundamental mode of the taper and then coupling back to the fundamental mode of slab #2. We assume that

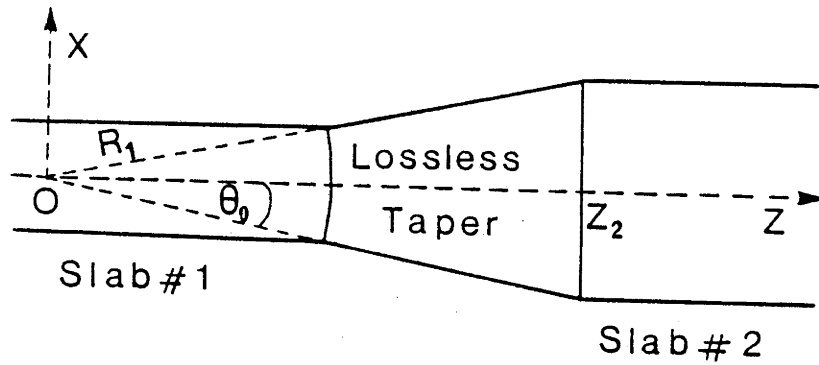


Fig. 4 Geometry of the mode-coupling calculation.

the taper is single-moded ($V_R < \pi/2$) and of sufficient length that any power coupled into higher order taper modes is radiated away, and a negligible fraction couples back into the fundamental mode of the second slab.

7.4.2 Mode Excitation

Let $\psi_1(X) \exp(i\beta_1 Z)$ and $\psi_2(X) \exp(i\beta_2 Z)$ denote the fundamental-mode fields of the two slabs, and $B_o(\theta) \mathcal{H}_\mu^{(1)}(kn_o R)$ ($\mu = i\sqrt{|v_o|}$) the fundamental-mode field of the taper.

To calculate the excitation of the taper mode by the incident slab field we use the orthogonality of the Marcatili modes over the circle $R = R_1$. Note that if the field on this circle is written $\Phi(\theta)$ and is outward propagating, then the fundamental mode of the taper is excited with modal amplitude

$$a_o = \frac{\int_{-\pi}^{\pi} \Phi(\theta) B_o(\theta) d\theta}{\mathcal{H}_\mu^{(1)}(kn_o R_1) \int_{-\pi}^{\pi} B_o^2(\theta) d\theta} \quad (21)$$

In our case the excitation is due to the slab mode, $\psi_1(X) \exp(i\beta_1 Z)$, and we can obtain $\Phi(\theta)$ by the change of coordinates $X = R \sin\theta$, $Z = R \cos\theta$. Provided this field subtends a small angle as seen from the origin (i.e. in terms of the spot-size, X_1 , we assume $X_1/R_1 \ll 1$) then we can take

$$\Phi(\theta) \approx \psi_1(R_1\theta) \exp[i\beta_1 R_1(1 - \theta^2/2)] \quad (22)$$

Here we have retained terms to second order in θ and thus take account of the wavefront curvature.

To calculate the excitation of the second slab we use the orthogonality of slab modes on the plane $Z = Z_2$. If the field on this plane is written $\tau(X)$, then the fundamental mode of the second slab is excited with modal amplitude

$$b_0 = \frac{\int_{-\infty}^{\infty} \tau(X) \psi_2(X) dX}{\exp(i\beta_2 Z_2) \int_{-\infty}^{\infty} \psi_2^2(X) dX} \quad (23)$$

Excitation of the second slab is due to the fundamental mode of the taper with modal amplitude a_0 . In any practical situation we will have $kn_0 R_1 \gg 1$, allowing us to replace the Hankel function of the taper field (and in Eqn (21)) by an asymptotic formula⁴. Thus, the field exciting the second slab is given by:

$$a_0 B_0(\theta) H_{\mu}^{(1)}(kn_0 R) \approx a_0 B_0(\theta) (2/\pi kn_0 R)^{1/2} \exp\{i[kn_0 R - \mu\pi/2 - \pi/4]\} \quad (24)$$

We obtain $\tau(X)$ from Eqn (24) by using the change of coordinates $R = (X^2 + Z^2)^{1/2}$ $\theta = \arctan(X/Z)$. Once again assuming that the taper field has angular spot-size $\theta_s \ll 1$, then we can take

$$\begin{aligned} \tau(X) = & a_0 B_0(X/Z_2) (2/\pi k n_0 Z_2)^{1/2} \\ & \times \exp\{i[kn_0(Z_2 + X^2/2Z_2) - \mu\pi/2 - \pi/4]\} \end{aligned} \quad (25)$$

Again we retain terms to second order in (X/Z_2) in the complex exponential to take account of wavefront curvature.

The coupling efficiency is defined by

$$\eta = P_2/P_1 \quad (26)$$

where P_1 is the incident power, P_2 the power coupled into the fundamental mode of the second slab. In scalar theory, Eqn (26) becomes

$$\eta = \frac{|b_0|^2 \int_{-\infty}^{\infty} \psi_2^2(x) dx}{\int_{-\infty}^{\infty} \psi_1^2(x) dx} \quad (27)$$

Obtaining b_0 from Eqns (21) - (23) and (25) and substituting into Eqn (27) gives

$$(Z_2/R_1) \eta =$$

$$\frac{\left| \int_{-\pi}^{\pi} \psi_1(R_1\theta) B_0(\theta) \exp(-is\theta^2) d\theta \right|^2 \left| \int_{-\infty}^{\infty} \psi_2(x) B_0(X/Z_2) \exp(iux^2) dx \right|^2}{\int_{-\infty}^{\infty} \psi_1^2(x) dx \int_{-\infty}^{\infty} \psi_2^2(x) dx \left| \int_{-\pi}^{\pi} B_0^2(\theta) d\theta \right|^2} \quad (28)$$

where

$$s = \beta_1 R_1 / 2 \quad (29a)$$

$$u = kn_0 / 2Z_2 \quad (29b)$$

7.4.3 Gaussian Approximation

We substitute Gaussian functions for the fields in Eqn (28), and extend the θ integrations to $\pm \infty$ (see the remarks following Eqn (15)). All the integrals can now be evaluated in closed form and we obtain for the coupling efficiency

$$\eta = \eta_{\text{spot}} \times \eta_{\text{curv}} \quad (30)$$

where

$$\eta_{\text{spot}} = \left[\frac{2(x_1/R_1\theta_s)}{1 + (x_1/R_1\theta_s)^2} \right] \left[\frac{2(x_2/Z_2\theta_s)}{1 + (x_2/Z_2\theta_s)^2} \right] \quad (31)$$

$$\eta_{\text{curv}} = \left[\frac{1}{1 + p^2} \right]^{1/2} \left[\frac{1}{1 + q^2} \right]^{1/2} \quad (32)$$

We have defined dimensionless parameters p and q by

$$\begin{aligned} p &= \beta_1 R_1 \theta_s^2 / \{1 + (R_1 \theta_s / X_1)^2\} \\ &= \alpha / \{1 + (R_1 \theta_s / X_1)^2\} \end{aligned} \quad (33a)$$

and

$$\begin{aligned}
 q &= kn_o z_2 \theta_s^2 / \{1 + (z_2 \theta_s / x_2)^2\} \\
 &= \bar{\alpha} / \{1 + (z_2 \theta_s / x_2)^2\}
 \end{aligned} \tag{33b}$$

These parameters are a combination of the spot-size ratios together with α and $\bar{\alpha}$, the significance of which we discuss in the next section. Note that

$$\begin{aligned}
 \alpha &= \beta_1 R_1 \theta_s^2 \approx kn_o R_1 \theta_s^2 \\
 &= (R_1 \theta_o / N_o) V_R f^2(V_R)
 \end{aligned} \tag{34a}$$

and

$$\begin{aligned}
 \bar{\alpha} &= kn_o z_2 \theta_s^2 \\
 &= (z_2 \theta_o / N_o) V_R f^2(V_R)
 \end{aligned} \tag{34b}$$

We note, in passing, that this problem has been reduced to the coupling of Gaussian beams, a problem considered in its full generality by Kogelnik⁷.

7.5 ANALYSING THE EFFICIENCY

The coupling efficiency is conveniently factored into two distinct components, η_{spot} and η_{curv} , and each of these can be further separated into contributions from either end of the taper. Comparison of Eqns (20) and (31) reveals that η_{spot} can be associated with the familiar loss due to mismatch of Gaussian spot-size at the slab/taper junctions, and depends only on the spot-size ratios. However, η_{curv} is a new term resulting from the mismatch of wavefront curvature, and depends on the spot-size ratios and the parameters α and $\bar{\alpha}$.

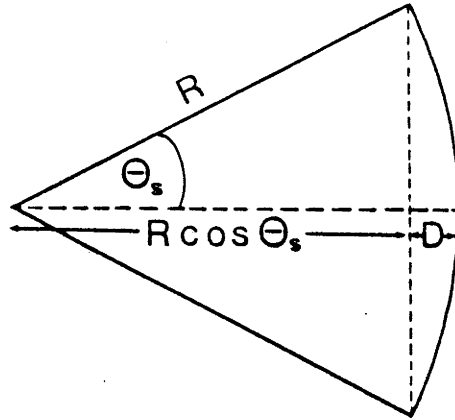


Fig. 5 Schematic diagram showing the curved wavefront of the fundamental mode of the taper (with angular spot-size θ_s) and planar wavefront of the fundamental mode of the slab. The maximum separation, labelled D , is given by $D = 2R \sin^2(\theta_s/2)$.

The physical significance of α and $\bar{\alpha}$ is that they give a measure of the maximum difference in phase between the curved (taper) and planar (slab) guided modes. This is illustrated in Fig. 5 which depicts schematically the curved wavefront of the fundamental mode of the taper, and the planar wavefront of the fundamental mode of the slab. If we set the phase difference on axis to zero (where the wavefronts touch in Fig. 5) then the maximum separation, labelled D in the figure, determines the maximum phase difference between the modal fields during excitation. In fact, as a rough approximation, the local propagation constant $\beta_R \approx kn_0$ (see Eqn (9) with R large enough so that $kn_0 R \gg |\nu|$) and $D \approx R \theta_s^2/2$, giving $\beta_R D \approx \alpha/2$ at the first slab/taper junction, and $\beta_R D \approx \bar{\alpha}/2$ at the second. Thus α and $\bar{\alpha}$ are critical parameters in determining the curvature loss at each slab/taper junction, and the larger α and $\bar{\alpha}$, the more significant is this loss.

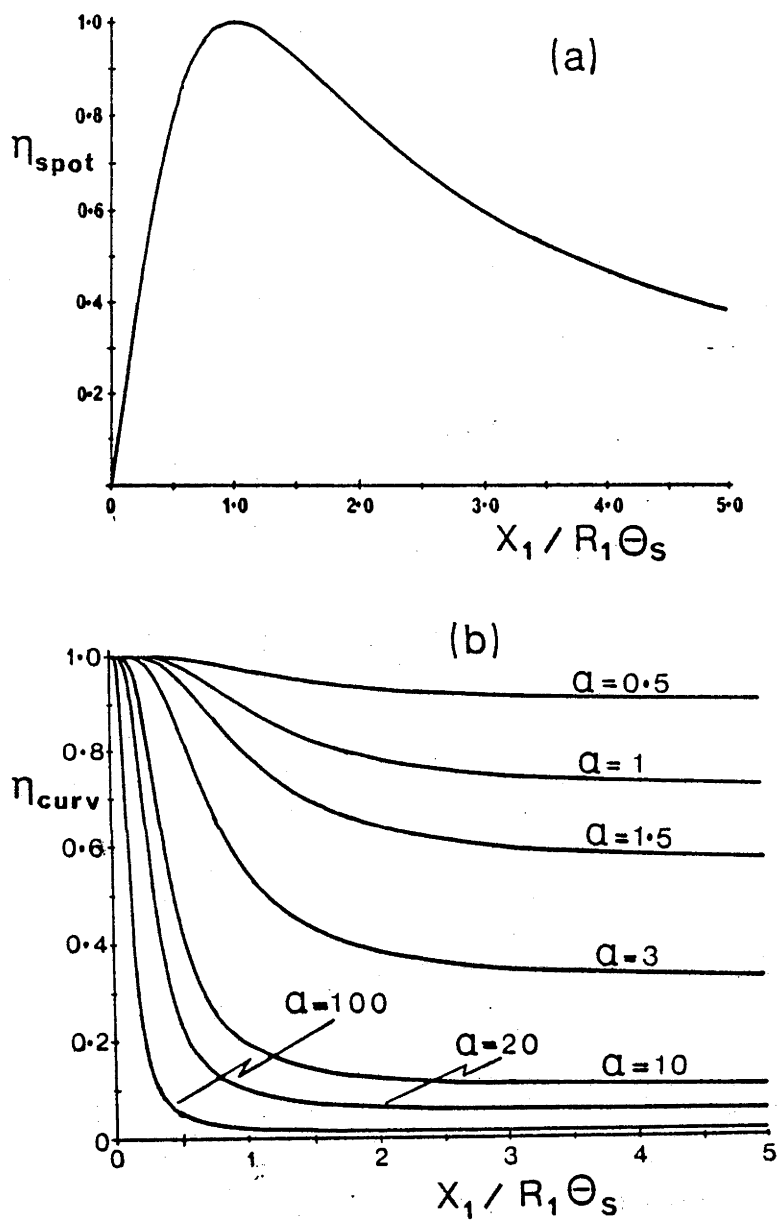


Fig. 6

Loss contributions at the first slab/taper junction due to mismatch of (a) spot-size, and (b) wavefront curvature. Note that in (b) the parameter $\alpha = \beta_1 R_1 \Theta_s^2$ is held fixed at the values indicated, and each curve approaches a horizontal asymptote located at $\eta_{\text{curv}} = 1/[1 + \alpha^2]^{1/2}$.

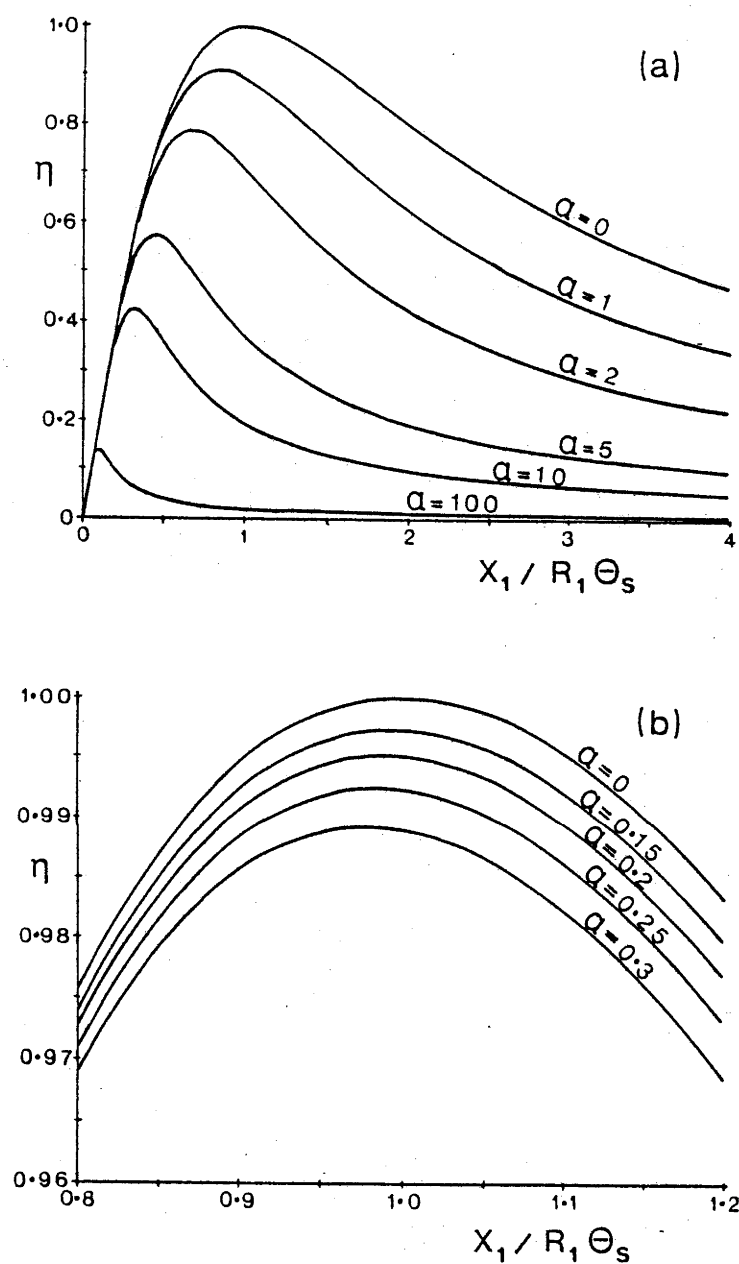


Fig. 7

Overall coupling efficiency at the first slab/taper junction, for various fixed values of α . In (a) the gross features are illustrated, while in (b) we show the detailed shift of the peak coupling efficiency for small values of α .

The two individual contributions to the coupling efficiency are shown graphically in Fig. 6. To simplify matters we only plot the first factor appearing in Eqns (31) - (32), corresponding to loss from the first slab/taper junction. Figure 6(a) shows η_{spot} versus $(X_1/R_1\theta_s)$ and demonstrates the critical dependence on the spot-size ratio (see Marcuse⁶). Figure 6(b) is a plot of η_{curv} versus $(X_1/R_1\theta_s)$ for various fixed values of α , and clearly illustrates the point that the curvature loss is less significant for smaller values of α . Note, from the figure, the strong dependence of η_{curv} on α when $\alpha \lesssim 10$ and the much weaker dependence when $\alpha \gtrsim 10$.

The overall maximum coupling efficiency (taking into consideration both factors, η_{spot} and η_{curv}) will occur approximately when $(X_1/R_1\theta_s) = 1$, but actually at a value somewhat less than unity due to the curvature effect. This is illustrated in Fig. 7(a), which depicts the overall coupling efficiency at the first slab/taper junction, as a function of $X_1/R_1\theta_s$, for various values of α . (Note that the same curves also give the overall coupling efficiency at the second slab/taper junction, if we replace $X_1/R_1\theta_s$ and α with $X_2/Z_2\theta_s$ and $\bar{\alpha}$.)

Finally, some comments regarding the utilization of such a taper in practical waveguide splicing. Waveguide splices with an insertion loss of 0.1 dB or less, are generally considered acceptable in long distance fiber optic communications. We use this figure as an example, and design the taper coupler so that the loss at either waveguide/taper junction is less than 0.05 dB (equivalent to a coupling efficiency of better than 98.9% at each junction). Figure 7(b) shows the peak coupling efficiency for small values of α . Clearly we require $\alpha \lesssim 0.3$ (and $\bar{\alpha} \lesssim 0.3$ at the second slab/taper junction) to accomplish this. A taper so designed has its peak

coupling efficiency when $X_1/R_1\theta_s$ (and $X_2/Z_2\theta_s$) is within 3% of unity (see Fig. 7(b)). With negligible error we can assume that the maximum coupling efficiency is obtained when $X_1/R_1\theta_s = 1$, $X_2/Z_2\theta_s = 1$ (i.e. spot-sizes matched, as described in section 7.3.1). This leads to the very simple expression for the dB insertion loss at the first slab/taper junction:

$$\begin{aligned} P &= -10 \log_{10} \eta \\ &= 5 \log_{10} \{1 + \alpha^2/4\} \end{aligned} \quad (35)$$

The insertion loss at the second slab/taper junction is obtained by replacing α with $\bar{\alpha}$ in the above equation. Equation (35) is plotted in Fig. (8).

Thus, a low loss taper splice (with insertion loss less than 0.1 dB) is designed with the phase parameters (Eqns (34)) satisfying $\alpha \leq 0.3$, $\bar{\alpha} \leq 0.3$ and spot-sizes matched (see Eqns (19)). The insertion loss for each junction is obtained from Fig. (8), and added together to obtain the total insertion loss. These equations show how to choose the taper parameters: if we specify the loss at each junction, then α and $\bar{\alpha}$ are fixed, and together with the spot-size matching conditions there are four equations which determine the four taper parameters: R_1, Z_2, θ_0 and N_0 . However, there may be non-optical constraints limiting the range of parameters that can be used. For example, the length of the device ($Z_2 - R_1$) may need to be controlled.

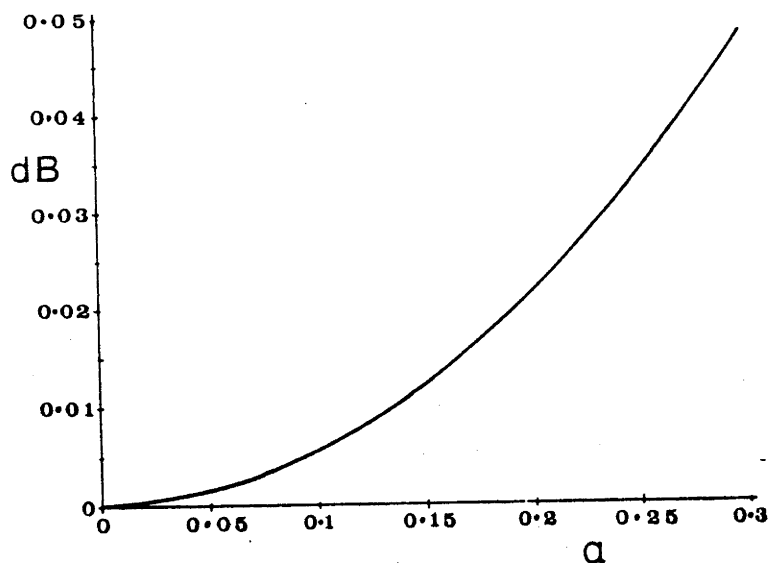


Fig. 8 Minimum insertion loss, in decibels, at a single slab/taper junction, as a function of α .

7.6 CONCLUSION

The Marcatili taper can be viewed as a mode expander for the step-index slab, for any slab mode. When used as a waveguide connector, losses occur at the slab/taper junction due to mismatch in spot-size (which can be eliminated through taper design) and wavefront curvature. These losses have been quantified in terms of the spot-sizes of the fundamental modes, and the phase parameters α and $\bar{\alpha}$.

In this chapter we have considered the practical situation of the taper arranged in physical contact with the slab waveguides. The results can easily be extended to cover the case in which the taper is separated from the slabs by a gap. Assuming Gaussian fields (coupling between fundamental modes) then the diffracted field from the end-face of slab #1 is simply a Gaussian beam. By spacing the slab and taper end-faces appropriately (but keeping the axes aligned), both the Gaussian beam spot-

size and wavefront curvature can be matched to the fundamental mode of the taper. Loss would now be dominated by higher order effects neglected in this paper (e.g. backscattering and Fresnel reflection loss, mismatch of cladding fields). At the other end of the taper, the taper field excites a second Gaussian beam with spot-size and wavefront curvature appropriate to that end-face. This second Gaussian beam cannot match the fundamental field of slab #2 because of the different wavefront curvatures, and the coupling efficiency can be obtained directly from Kogelnik⁷.

Finally we observe that it may be feasible to use a thin lens at the slab/taper junctions (perhaps formed by the curved end-face of the taper itself, with fluid of appropriate refractive index in between) to transform the planar into curved wavefronts, and vice-versa.

7.7 REFERENCES

1. A.W. Snyder and J.D. Love, "Optical Waveguide Theory," (Chapman and Hall, London, 1983).
2. E.-G. Neumann, "Beam Transformers for obtaining Low-Loss Splices between Dissimilar Single-Mode Fibers," J. Opt. Soc. Am. A4, 1021 (1987).
3. E.A.J. Marcatili, "Dielectric Tapers with Curved Axes and No Loss," IEEE J. Quantum Electron. QE-21, 307 (1985).
4. M. Abramowitz and I.A. Stegun, "Handbook of Mathematical Functions," (Dover, New York, 1972).

5. D. Marcuse, "Light Transmission Optics," (Van Nostrand Reinhold, New York, 1982).
6. D. Marcuse, "Loss Analysis of Single-Mode Fiber Splices," Bell Syst. Tech. J. 56, 703 (1977).
7. H. Kogelnik, "Coupling and Conversion Coefficients for Optical Modes," in Quasi-Optics: Microwave Research Institute Symposia Series, 14, (Polytechnic Press, New York, 1964).

CHAPTER 8

AN APPLICATION IN VISION: A NEW MODEL FOR ANALYSING CONE ELEMENTS IN VISUAL SYSTEMS

8.1 INTRODUCTION

Invertebrate eyes show unusual diversity in structure, reflecting a rich collection of optical systems of fundamentally different design (see, for example, the review by Land¹). Individual optical components can usually be identified as lenses, lens cylinders², waveguides³, reflectors⁴ and light concentrators⁵. Recently, Pask and Bertilone⁶ introduced a new optical element which behaves as a combination of lens cylinder, waveguide and light concentrator, which may be used to model cone shaped graded-index elements in visual systems. In particular, it explains the recent, distinctive measurements on the crystalline cone of Butterfly eyes, and leads to a new understanding of the optical principles underlying its design.

In this chapter we give details of the cone model, and describe its elementary optical properties. We discuss the physical arguments supporting its likely application in vision, and demonstrate how it closely models the behavior of the crystalline cone of the Butterfly eye, recently studied by Nilsson, Land and Howard⁷⁻⁸.

8.2 THE CONE MODEL

The cone, with half-angle θ_0 , is depicted in Fig. 1(a). We define spherical polar (R, θ, ϕ) and cylindrical (r, z, ϕ) coordinates, where ϕ is the azimuthal angle. The refractive index distribution inside the cone is taken to be

$$n^2 = n_0^2 \{1 + 2\Delta(\theta_1^2 - \theta^2)/R^2\} \quad (1)$$

where n_0 , Δ and θ_1 are constants. We set $\theta_1 > \theta_0$ (if $\theta_1 = \theta_0$, then the cone edge has constant index n_0).

When the cone angle is small, $\theta_0 \ll 1$, (which is certainly the case in most vision applications) we can make the approximate change of coordinates $R \approx z$, $\theta \approx r/z$ to give

$$n^2 \approx n_0^2 \left[1 + \left(2\Delta\theta_1^2/z^2\right)\right] - \left(2\Delta n_0^2/z^4\right) r^2 \quad (2)$$

The above equation describes a tapered quadratic-index rod. As the cone narrows, the characteristic radius of the transverse grading decreases, and the on-axis refractive index increases. Contours of constant refractive index are plotted schematically in Fig. 1(b).

The refractive index outside the cone is taken to be uniform, n_s . When $n_s = n_0$ the system is separable, and the rays and fields can be analysed using the formalism developed in chapters 2 and 3. Guided wave (bound ray) solutions are found, for which light propagates down the taper without radiation. Lowering the external refractive index is not expected to change this behavior (certainly the bound ray solutions will be unaffected).

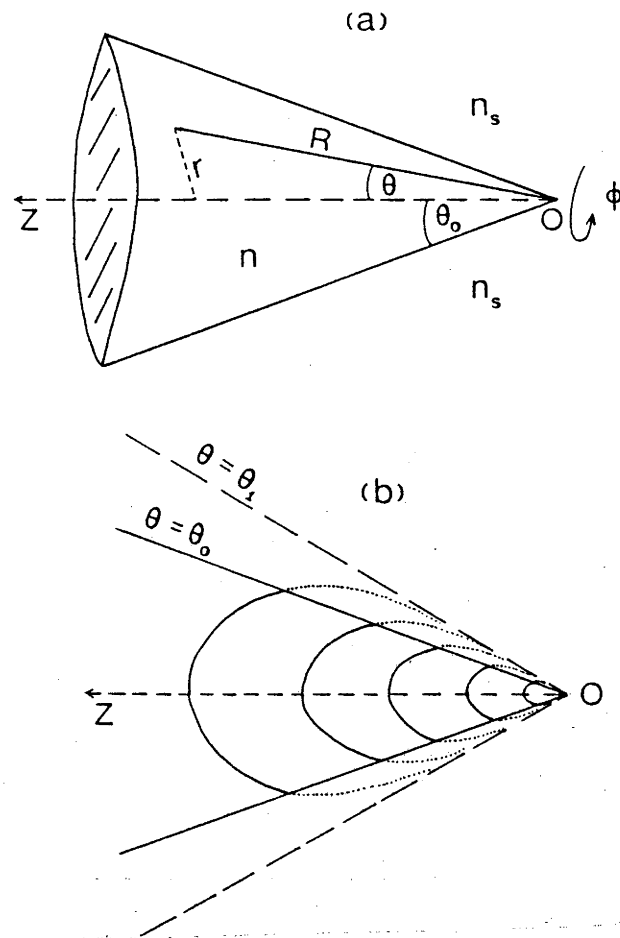


Fig. 1 (a) Coordinate systems. A point in the cone can be described by spherical polar (R, θ, ϕ) or cylindrical (r, z, ϕ) coordinates.
 (b) Schematic depiction of the contours of constant refractive index in a meridional plane.

It should be mentioned that Sharma and Goyal⁹ obtained the small-angle form of the index distribution, Eq. (2), from constant V (waveguide parameter¹⁰) considerations. Without recognizing the very special guiding properties of such a taper, they plotted ray paths by numerically integrating the ray equation. Interestingly, they suggested that it might be useful for modelling the cone receptors in the human retina, but this is highly speculative. The cone receptors are so small that conventional measurement techniques are unable to resolve their detailed optical structure, and there is little evidence that they are graded-index elements.

8.3 MERIDIONAL RAYS IN THE CONE

We begin with the separable case, $n_s = n_o$. Following chapter 2, rays are characterized by the ray invariant κ , Eq. (2.20). The ray path equation is given by Eq. (2.24a) with $G(\theta)$ defined by

$$G(\theta) = \begin{cases} \theta^2 - \theta_1^2 & , \quad |\theta| < \theta_o \\ 0 & , \quad \theta_o < |\theta| < \pi \end{cases} \quad (3)$$

Of interest here are the meridional rays, which have skew invariant $w = 0$ (see Eq. (2.15)), and are confined to a plane passing through the cone axis. In this situation it is useful to interpret (R, θ) as plane polar coordinates, and to allow θ to take on negative values. The r variable now labels the axis perpendicular to the cone axis (in the meridional plane) and can also assume negative values.

It is convenient to label the rays using the dimensionless quantity μ , which is related to the ray invariant κ by

$$\mu = \sqrt{(\kappa + 2\Delta n_o^2 \theta_1^2) / 2\Delta n_o^2} \quad (4)$$

From the ray path equation it follows that rays with $0 < \mu < \theta_o$ are bound, and confined by caustics to oscillate within the limits $|\theta| < \mu$. Rays with $\theta_o < \mu < \theta_1$ are also bound, but reflect off the cone edges: $|\theta| < \theta_o$. Rays with $\mu > \theta_1$ are refracting.

The bound ray solutions are obtained by solving

$$\frac{d\theta}{dR} = \pm \frac{\sqrt{2\Delta} \sqrt{(\mu^2 - \theta^2)}}{R\sqrt{(\sigma^2 + R^2)}} \quad (5a)$$

where σ is a constant, defined by

$$\sigma = \sqrt{2\Delta} \sqrt{(\theta_1^2 - \mu^2)} \quad (5b)$$

The bound ray solutions are shown schematically in Fig. 2. Integrating Eq. (5a), we obtain a complete, explicit formula for those rays confined by caustics inside the cone ($0 < \mu < \theta_0$) :

$$\begin{aligned} \theta = \mu \sin \{ \pm (\sqrt{2\Delta}/\sigma) [\sinh^{-1}(\sigma/R) - \sinh^{-1}(\sigma/R_i)] \\ + \sin^{-1}(\theta_i/\mu) \} \end{aligned} \quad (6a)$$

where (R_i, θ_i) is any point on the trajectory, and we use $\pm$ if $d\theta/dR < 0$ or $d\theta/dR > 0$ at this point. Alternatively, we can write

$$\theta = \mu \cos \{ (\sqrt{2\Delta}/\sigma) [\sinh^{-1}(\sigma/R) - \sinh^{-1}(\sigma/R_0)] \} \quad (6b)$$

where R_0 is any value of R at which the ray touches the upper caustic: $\theta(R_0) = \mu$ (see Fig. 2(a)). Note that the period of the ray oscillations rapidly diminishes as the cone apex is approached.

The other class of bound rays ($\theta_0 < \mu < \theta_1$) is of less interest (see later). The formula for these rays is

$$\theta = \mu \sin \{ \pm (\sqrt{2\Delta}/\sigma) \sinh^{-1}(\sigma/R) + \text{CONST.} \} \quad (7)$$

where the value of the constant will change following each reflection off the cone edge.

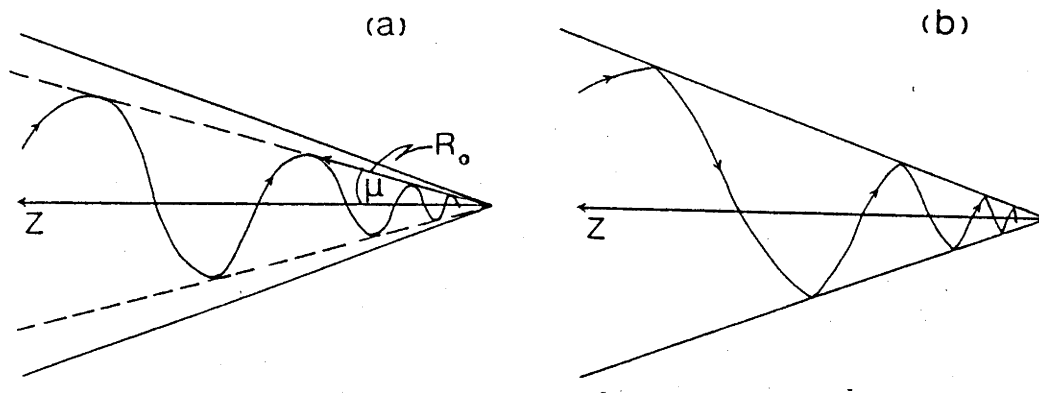


Fig. 2 Schematic illustration of the bound meridional rays with
(a) $0 < \mu < \theta_0$ and (b) $\theta_0 < \mu < \theta_1$.

8.3.1 WEAK GRADING ASSUMPTION

In biological applications, refractive index gradients are always small. In the crystalline cone of the Butterfly, for example, the index gradient is due to an inhomogeneous protein concentration, and is limited to very small variation. (Currently used fabrication techniques also limit manufactured index gradients to small values.) Thus we can assume

$$\frac{n^2 - n_0^2}{n_0^2} = \frac{2\Delta(\theta_1^2 - \theta^2)}{R^2} \ll 1 \quad (8)$$

(see Eq. (1)) will hold at every point inside the region of cone which is of interest. The refractive index inside the cone can now be written approximately as

$$n \approx n_0 + n_0 \Delta(\theta_1^2 - \theta^2)/R^2 \quad (9)$$

Within this assumption, $\sigma/R \ll 1$ and we can expand the inverse hyperbolic functions of Eqs. (6) and (7) to first order in σ/R , with negligible error. Thus, Eq. 6(b) reduces to the remarkably simple expression

$$\theta = \mu \cos \left\{ \sqrt{2\Delta} \left[\left(1/R \right) - \left(1/R_0 \right) \right] \right\} \quad (10)$$

All of the above rays are periodically focused inside the cone, at locations on the axis given by

$$z_f = R_0 / \left\{ 1 + (2m + 1)(\pi/2)(R_0/\sqrt{2\Delta}) \right\} \quad (11)$$

where m is an integer.

8.4 SPECIAL PROPERTIES OF THE CONE

The cone is an unusual optical element which behaves as a combination waveguide, lens and light concentrator.

The guidance of light was briefly discussed with reference to the bound rays. The important point is that even within wave theory there will be a finite number of bound modal fields (with curved wavefronts) guided, without loss, by the cone (see chapter 3 or Marcatili¹¹).

The internal self-focusing of rays inside the cone has been pointed out (previous section). In fact, the quadratic variation of refractive index in the transverse plane ensures that any transverse slice of the cone will behave like a lens, and will form images of paraxial objects (see Gomez-Reino et al.¹²).

The cone also behaves as a light collector which can be used to effectively increase the light capture area of a detector placed at its narrow end. Bound rays launched into the cone remain trapped, and the conical geometry compresses them into a progressively smaller region. Figure 3 shows schematically the spread of ray directions (meridional rays only) which can be launched into the curved endface of the cone, and be captured as bound rays:

$$\sin \alpha_m(\theta_i) = (n_o/n_b)(\sqrt{2\Delta}/R_e) \sqrt{(\theta_1^2 - \theta_i^2)} \quad (12)$$

Here R_e defines the cone endface, and θ_i determines the point on the endface into which rays are directed. The refractive index outside the cone is n_b , and α_m labels the maximum launch angle (relative to the radial direction - see Fig. 3). The spread of rays collected by the cone is maximum on the cone axis ($\theta_i = 0$) and decreases away from the axis:

$$\sin \alpha_m(\theta_i)/\sin \alpha_m(0) = \sqrt{[1 - (\theta_i/\theta_1)^2]} \quad (13)$$

The configuration shown in Fig. 4 is of most interest in the vision application. A detector (photoreceptor) is joined to the narrow end of the cone, and light is directed into the cone by an external lens. The lens need not form an image on the cone endface, but must launch bound rays efficiently. In wave theory, the Fresnel (near-field) of the lens excites bound modes of the cone, which propagate down to the detector without loss. The detector, in animal eyes, is a photoreceptor which behaves like an absorbing waveguide³. In our model, the cone/photoreceptor may be viewed as an extended waveguide with a large light-capture area. This is a

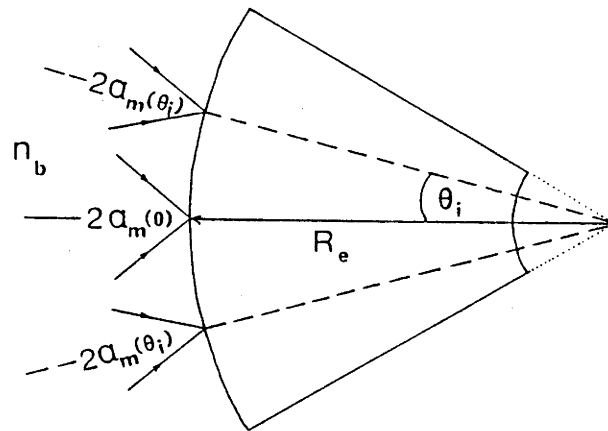


Fig. 3 Bound (meridional) rays launched into the curved endface of the cone. The spread of ray angles, at the point (R_i, θ_i) is denoted $\alpha_m(\theta_i)$.

highly attractive system, particularly when compared with the same configuration using a homogeneous index taper - which radiates continuously and is a relatively poor light collector.

Graded-Index cone elements are found in some invertebrate eyes (see, for example, section 8.6 on the Butterfly compound eye) in the configuration depicted in Fig. 4. The attractive physical features of our model make it (or a simple variation) a likely candidate for modelling their behavior. (Note that guidance will occur with more general grading functions, $G(\theta)$, instead of $(\theta_1^2 - \theta^2)$ in Eq. (1), as discussed in chapters 2 and 3. However, the latter is the simplest and appears to be appropriate for vision studies.) Verification of the model must come from experiment, and the simplest experiment which can be carried out is to slice the cone into thin transverse sections, and measure the focusing power (reciprocal of focal length) of each slice. In the next section we derive a formula for the power per unit length of cone slices.

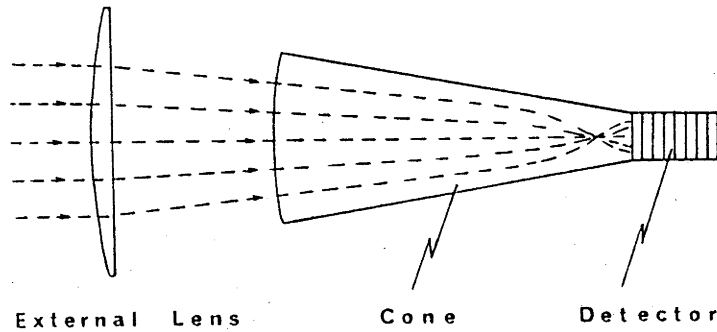


Fig. 4 Light collected by an external lens and concentrated down to the detector, by the cone.

8.5 FOCUSING POWER OF CONE SLICES

In this section we shall derive a formula for the focusing power per unit length, of a transverse section of cone with mid-point at $\bar{z}$ and width L : denoted $P(\bar{z}, L)/L$.

Figure 5 depicts a transverse slice of the cone, of width L , with flat endfaces at z_i and $z_i - L$. The focal length, f , is obtained by considering rays incident normal to the endface at z_i (corresponding to a point object at infinity). Outside the cone section, the refractive index is assumed to be n_d (in practice, the experiment would usually take place in air, with $n_d = 1$).

A typical ray is shown schematically in Fig. 5. If the initial angle, θ_i , satisfies

$$|\theta_i| < \theta \quad (14a)$$

where θ is the solution ($0 < \theta < \pi/2$) of the transcendental equation

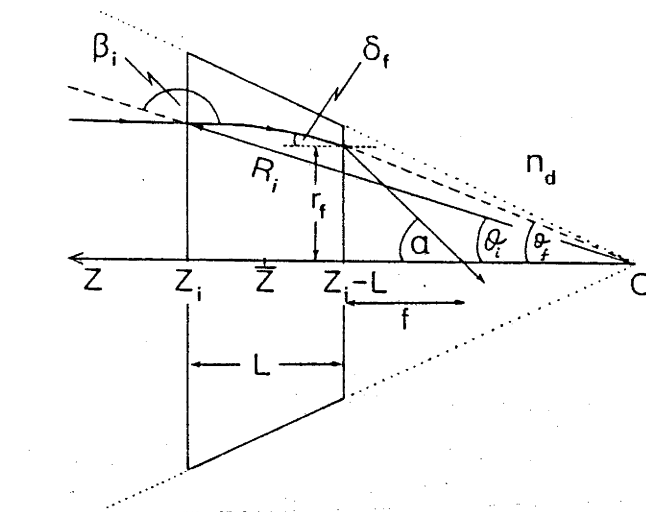


Fig. 5 Transverse section of cone, of width L , showing quantities used in power derivation.

$$\theta_1^2 \sin^2 \theta + \theta^2 \cos^2 \theta + (z_1^2 / 2\Delta) \tan^2 \theta = \theta_0^2 \quad (14b)$$

then the ray in the cone section will be bound, of the type described by Eq. (6). (This follows from Eq. (4) and the ray invariant Eq. (2.20).) In fact, for small θ_0 ,

$$\theta \approx \theta_0 / \sqrt{1 + \theta_1^2 + z_1^2 / 2\Delta} \quad (14c)$$

Thus all the incident rays, with θ_1 small enough, will be of the type depicted in Fig. 2(a).

The constants μ and σ , for the ray in Fig. 5, are obtained using the initial conditions:

$$R_i = z_i / \cos \theta_i \quad (15a)$$

$$\beta_i = \pi - \theta_i \quad (15b)$$

where β_i is the angle between the ray path and the radial direction (see Fig. 3 of chapter 2). Under the assumptions $\theta_i \ll 1$ (small cone-angle) and $R_i^2/2\Delta \gg \theta_1^2$ (weak index-grading) we find to second order in θ_i :

$$\mu \approx \theta_i \sqrt{1 + z_i^2/2\Delta} \quad (16a)$$

$$\sigma \approx \sqrt{2\Delta} \sqrt{[\theta_1^2 - \theta_i^2(1 + z_i^2/2\Delta)]} \quad (16b)$$

Substituting Eqs. (16) into the ray formula, Eq. (6a), gives the explicit ray path in the cone section. The weak grading assumption allows us to expand the inverse hyperbolics to first order (see section 8.3.1) to give

$$\theta \approx [\theta_i \sqrt{1 + z_i^2/2\Delta}] \sin\{(\sqrt{2\Delta}/R) - (\sqrt{2\Delta}/R_i) + \sin^{-1}[1/\sqrt{1 + z_i^2/2\Delta}]\} \quad (17)$$

To summarize the assumptions we shall use in this analysis:

1. WEAK GRADING: $2\Delta(\theta_1^2 - \theta^2)/R^2 \ll 1$
(at every point in the cone section).
2. SMALL CONE-ANGLE: $\theta_o \sqrt{1 + z_i^2/2\Delta} \ll 1$
(thus ensuring $|\theta| \ll 1$, see Eq. (17)).

$$3. \quad \text{PARAXIAL RAYS: } \frac{\theta_o \sqrt{[(1 + z_i^2/2\Delta)(1 + (z_i - L)^2/2\Delta)]}}{(z_i - L)/\sqrt{2\Delta}} \ll 1$$

(this ensures that all the rays will exit the cone section paraxially - see later).

Assumption 2 allows us to convert to cylindrical coordinates, using the approximate transformation $R \approx z$, $\theta \approx r/z$:

$$r \approx [\theta_i \sqrt{(1 + z_i^2/2\Delta)}] \sin\{(\sqrt{2\Delta}/z) - (\sqrt{2\Delta}/z_i) + \sin^{-1}[1/\sqrt{(1 + z_i^2/2\Delta)}]\} \quad (18)$$

Differentiating Eq. (18) with respect to z , gives the slope of the ray path:

$$\begin{aligned} dr/dz &= \tan\delta \\ &= \{(\theta_i \sqrt{2\Delta}/z) \sqrt{[(1 + z_i^2/2\Delta)(1 + z^2/2\Delta)]}\} \sin\{(\sqrt{2\Delta}/z) - (\sqrt{2\Delta}/z_i) \\ &\quad + \sin^{-1}[1/\sqrt{(1 + z_i^2/2\Delta)}] - \tan^{-1}[\sqrt{2\Delta}/z]\} \end{aligned} \quad (19)$$

where δ is the angle between the ray path and the cone axis (Fig. 5). Clearly, Assumption 3 ensures that each ray will leave the cone slice paraxially ($\delta \ll 1$).

From Fig. 5, the power of the cone section is given by

$$1/f = (\tan\alpha)/r_f \quad (20)$$

where α is the angle at which the ray emerges from the slice, and r_f is its distance from the cone axis. From the assumption that the ray exits paraxially, we can use the paraxial approximation of Snell's law to write -

$$1/f \approx (n_f \tan \delta_f) / (n_d r_f) \quad (21)$$

where n_f and $\tan \delta_f$ refer to the index and ray slope at the point where the ray leaves the cone (Fig. 5). We immediately obtain the required power formula by substituting $r_f = r(z_i - L)$ (from Eq. (18)), $\tan \delta_f = \tan \delta(z_i - L)$ (from Eq. (19)), and $n_f \approx n_o$ (weak grading assumption) into Eq. (21). We obtain

$$\frac{1}{f} \approx \frac{n_o L}{n_d (z_i - L)^2} \left\{ \frac{(\sqrt{2\Delta}/L) [1 + z_i (z_i - L)/2\Delta] \tan[\sqrt{2\Delta} L/z_i (z_i - L)] - 1}{(z_i/\sqrt{2\Delta}) \tan[\sqrt{2\Delta} L/z_i (z_i - L)] + 1} \right\} \quad (22)$$

It is more useful to express the power of the cone section in terms of the midpoint of the section: $\bar{z} = z_i - L/2$. The power per unit length ($1/Lf$) can now be written

$$[P(\bar{z}, L)/L] \equiv (1/fL)$$

$$= \frac{n_o}{n_d} \frac{1}{[\bar{z} - L/2]^2} \left\{ \frac{[1 + 2\Delta/(\bar{z}^2 - L^2/4)] (\tanh \eta)/\eta - 1}{[L/(\bar{z} - L/2)] (\tanh \eta)/\eta + 1} \right\} \quad (23a)$$

where

$$\eta = \sqrt{2\Delta} L / (\bar{z}^2 - L^2/4) \quad (23b)$$

The power per unit length, $P(\bar{z}, L)/L$, depends strongly on the width, L , of the section of cone. In the limit $L \rightarrow 0$, representing infinitely thin sections, the power per unit length reduces to a very simple formula:

$$\left[\frac{P(z, L)}{L} \right]_{L \rightarrow 0} = \frac{2\Delta n_o}{n_d z^4} \quad (24)$$

(Here we use z to denote the location of the section, with no confusion.) Equation (24) depends solely on the cone parameters (and n_d) and may be viewed as characterizing the cone. Figure 6 shows a schematic plot of Eq. (24). In any real experiment, power measurements will be made on sections of non-zero thickness, and the resulting curve for power per unit length will lie above the idealized $L \rightarrow 0$ curve (see the broken curve in Fig. 6). (This assumes that the section is thin enough to exclude multiple turning points in the ray path!)

To conclude this section, it is worthwhile comparing the above results with those obtained for a cylindrical quadratic-index rod (lens cylinder) with index distribution

$$n^2 = a - br^2 \quad (25)$$

where a and b are constant. The focal length of such a rod (of length L) is given by Marcuse¹³ but must be modified to correct for refraction of the ray as it exits the endface. For paraxial rays, this correction factor is $n_d/\sqrt{a}$ (n_d is the index outside the rod) and we obtain:

$$\left[\frac{1}{fL} \right] = \frac{\sqrt{b}}{n_d L} \tan[L \sqrt{(b/a)}] \quad (26)$$

For an infinitely thin section of lens cylinder, Eq. (26) reduces to

$$\left[\frac{1}{fL} \right]_{L \rightarrow 0} = \frac{b}{n_d \sqrt{a}} \quad (27)$$

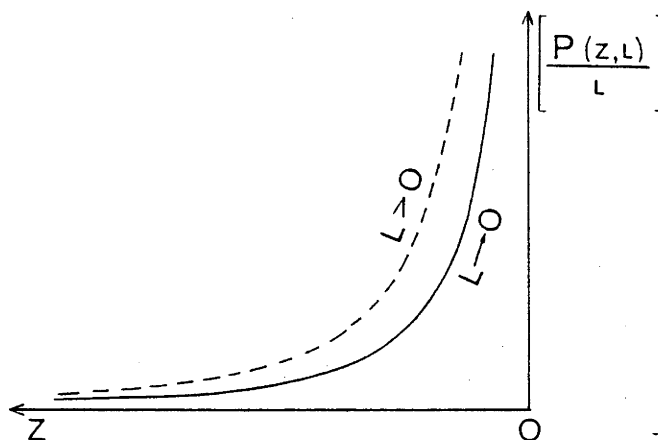


Fig. 6 Schematic diagram of the power per unit length for infinitely thin cone sections (solid line), and slices of finite thickness (dashed line).

Note from Eq. (2) that the local lens cylinder which matches the index distribution of the cone in the transverse plane at z , has parameters $a = n_o^2(1 + 2\Delta\theta_1^2/z^2) \approx n_o^2$ (weak grading) and $b = 2\Delta n_o^2/z^4$. Substituting into Eq. (27) shows that for infinitely thin slices, the power per unit length for the cone and local lens cylinder are the same. This is to be expected because the tapering vanishes in an infinitely thin slice of cone, and gives a useful verification of the power formula.

8.6 COMPOUND EYE OF THE BUTTERFLY

A recent study of a Butterfly compound eye has been published by Nilsson, Land and Howard⁷⁻⁸. Their experiments involved a common Australian Butterfly - *Heteronympha merope*. Figure 7 is a schematic diagram illustrating the structure of a single ommatidium (i.e. a single unit of the compound eye). The cornea is a thick lens which directs rays from a distant object to a point in the vicinity of the photoreceptor endface. A prominent feature is a comparatively large crystalline cone

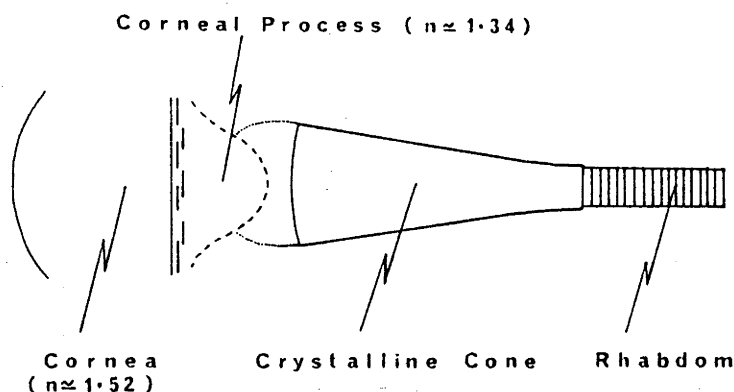


Fig. 7 Schematic illustration of the structure of a Butterfly ommatidium. Note that the corneal process provides a jump in refractive index from the cornea, giving it a thick lens behavior. The pigment cells, which surround the crystalline cone, are not shown.

(about $12.2\ \mu\text{m}$ in width at the wide end, about $2.6\ \mu\text{m}$ at the narrow end, and around $39\ \mu\text{m}$ in length) which is attached to the rhabdom (photoreceptor). Note from the dimensions of the cone, that light is brought from a many-wavelengths size region at the wide (distal) end, to a few-wavelengths size region at the narrow (proximal) end. This configuration is analogous to Fig. 4.

The Butterfly eye is known to have apposition optics (there is no clear-zone, see chapter 1) and such eyes are typically found to be simple focusing systems, with the cornea focusing light from a distant object directly onto the photoreceptor endface. Consequently, Nilsson and coworkers expected to find the crystalline cone optically homogeneous, serving only to physically separate the photoreceptor from the cornea. Due to the tiny dimensions of the cone, particularly at its narrow end, direction measurement techniques (such as interference microscopy) are

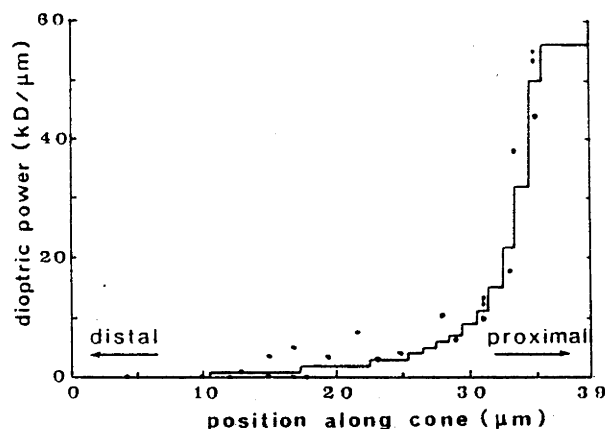


Fig. 8 Experimental power measurements from slices of crystalline cone 4-5 μm in width (reference 8).

inadequate for resolving the optical structure. Nevertheless, they performed the simple experiment of slicing the cone into thin transverse sections (4-5 μm in width⁸) and measuring the focusing power per unit length. They were very surprised to obtain the result shown in Fig. 8. The steeply rising curve is indicative of a strongly tapered GRIN rod. The same kind of behavior was predicted with our cone model (section 8.5), and because of the attractive physical properties of our model, we suggest that it (or a simple variation) may correctly describe the cone behavior.

In fitting our cone model to the data, we took account of the observation by Nilsson that the final 4 μm of the crystalline cone (at the proximal end) is cylindrical in appearance. They suggest that it may be an untapered lens cylinder. Presumably there would be a transition region near the cone/photoreceptor junction, so that the curved wavefronts of the taper modes can be matched to the planar wavefronts of the photoreceptor waveguide-modes. (It might be suggested that the index distribution in the cone may be adiabatically untapered, at the proximal end, to flatten the curved wavefronts.) Consequently, it is very reasonable to suggest that the final few microns of the cone is essentially untapered. We have

modelled the final 4 μm of crystalline cone as a lens cylinder, joined abruptly to the cone, with refractive index matching that of the end of the cone:

$$n^2 \approx n_o^2 - (2\Delta n_o^2/z_1^4) r^2 \quad (28)$$

(see Eq. (2) with the weak guidance assumption). Here z_1 denotes the end-face of the cone, and the start of the lens cylinder (see Fig. 9). In reality, we would expect a gradual change from cone to lens cylinder, but this would be difficult to model and the details of the transition are unknown.

8.7 FITTING MODEL TO BUTTERFLY DATA

Our model of the Butterfly crystalline cone is depicted schematically in Fig. 9. We now proceed to determine the model parameters - n_o , Δ and θ_1 .

Firstly, the location of the cone apex. From the measured dimensions of the cone, simple geometry gives $z_1 \approx 9.5 \mu\text{m}$ and the radius of the outer (distal) endface of the cone $R_e \approx 44.5 \mu\text{m}$ (see Fig. 9).

From measurements by Nilsson and coworkers, the refractive index at the distal end of the cone is known to be about 1.4, with very slight grading. Consequently, we set $n_o = 1.4$ in our model.

The final 4 μm of the Butterfly crystalline cone (which we model as a lens cylinder) has a measured power per unit length of around 55 kD/ μm . Equations (26) and (28) give for this section:

$$(1/fL) = (\sqrt{2\Delta}n_o/n_d L z_1^2) \tan(\sqrt{2\Delta}L/z_1^2) \times 10^3 \text{ kD}/\mu\text{m} \quad (29)$$

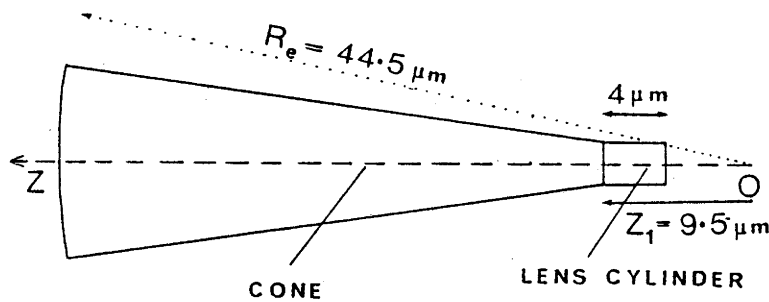


Fig. 9 Butterfly cone model.

(the factor 10^3 brings the units to kiloDiopters per micron). Setting $(1/fL) = 55 \text{ kD}/\mu\text{m}$, $L = 4 \mu\text{m}$, $z_1 = 9.5 \mu\text{m}$, $n_o = 1.4$ and $n_d = 1$ (Nilsson's measurements were made in air) we deduce $2\Delta \approx 263 \text{ sq.}\mu\text{m}$.

The parameters $n_o (\approx 1.4)$ and $2\Delta (\approx 263 \text{ sq.}\mu\text{m})$ determine the power per unit length for the whole crystalline cone. For infinitely thin sections, the power per unit length is plotted in Fig. 10 (dashed line), and this curve characterizes the model (note the constant section at the proximal end of the model, corresponding to the lens cylinder). The experiment of Nilsson, Land and Howard was carried out with cone slices $4\text{--}5 \mu\text{m}$ in width⁸. Figure 10 also shows the power per unit length predicted from our model, when $L = 4 \mu\text{m}$ (solid line). This curve is defined by Eq. (23) right up until $\bar{z}$ (the midpoint of the section) is $2 \mu\text{m}$ from the start of the lens cylinder. The $4 \mu\text{m}$ sections will subsequently have both cone and lens cylinder parts, and the power will be

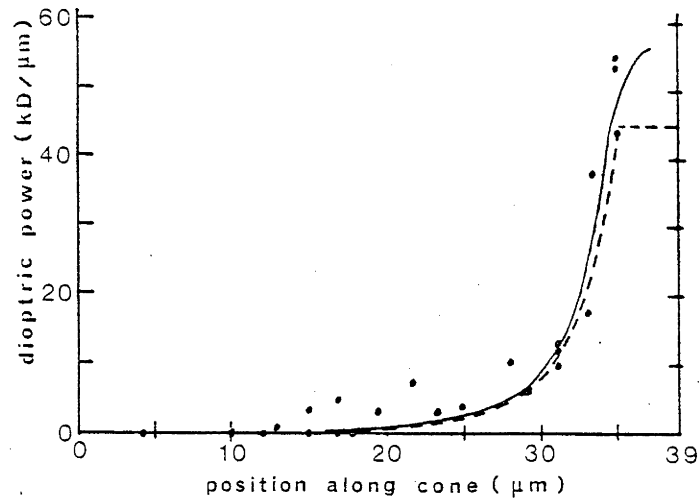


Fig. 10 Power per unit length predicted from the cone model. Dashed line is for infinitely thin slices, solid line is for slices of thickness $4 \mu\text{m}$. The data points are taken from reference 8.

some average of the two. When $\bar{z}$ reaches the middle of the lens cylinder, the section will be all lens cylinder, with power given by Eq. (26). The predictions of the model can be compared with experimental measurements (data points in Fig. 10) and the agreement appears to be quite good. This is an encouraging verification of the model's validity.

The final parameter in our model, θ_1 , remains to be determined. (Note that the ray paths in the weakly graded cone (Eq. (10)) depend negligibly on this parameter, and consequently it does not appear in the power formula.) The geometry of Fig. 9 gives a cone angle $\theta_0 \approx 0.136$ and we require $\theta_1 > \theta_0$. If θ_1 is too large, however, then the refractive index inside the cone becomes unrealistically big. For example, from Eq. (9) the maximum refractive index inside the cone is given by

$$\begin{aligned}
 n_{\text{axis}}(z_1) &= 1.4 + (2.04) \theta_1^2 \\
 &= 1.45 \quad \text{for } \theta_1 = 0.16 \\
 &= 1.48 \quad \text{for } \theta_1 = 0.20 \\
 &= 1.52 \quad \text{for } \theta_1 = 0.24
 \end{aligned}$$

Increasing θ_1 much further gives index values which are never observed in insect eyes. Thus, the refractive index limits θ_1 to comparatively small values. Conversely, a larger θ_1 gives a greater spread of ray angles captured by the cone (see Eq. (12)) and this determines the angular sensitivity of the system. The actual value of θ_1 will be a compromise between the two.

Before continuing, we check our parameter values with the assumptions used in section 8.5 (including weak grading).

8.7.1 CHECK ON PARAMETER VALUES

In checking the assumptions of section 8.5, we give numerical examples which deal with worst possible cases. The actual error will often be much less than is suggested here.

1. WEAK GRADING ASSUMPTION

$$\begin{aligned}
 [2\Delta(\theta_1^2 - \theta^2)/R^2]_{\text{max}} &= 2\Delta\theta_1^2/z_1^2 \\
 &= 0.17 \quad \text{for } \theta_1 = 0.24
 \end{aligned}$$

This assumption is certainly valid for the range of θ_1 of interest here. For example, $\sinh^{-1}(0.17) = 0.169$, and the approximation $\sinh^{-1}(Y) \approx Y$ in the ray formula will be very accurate.

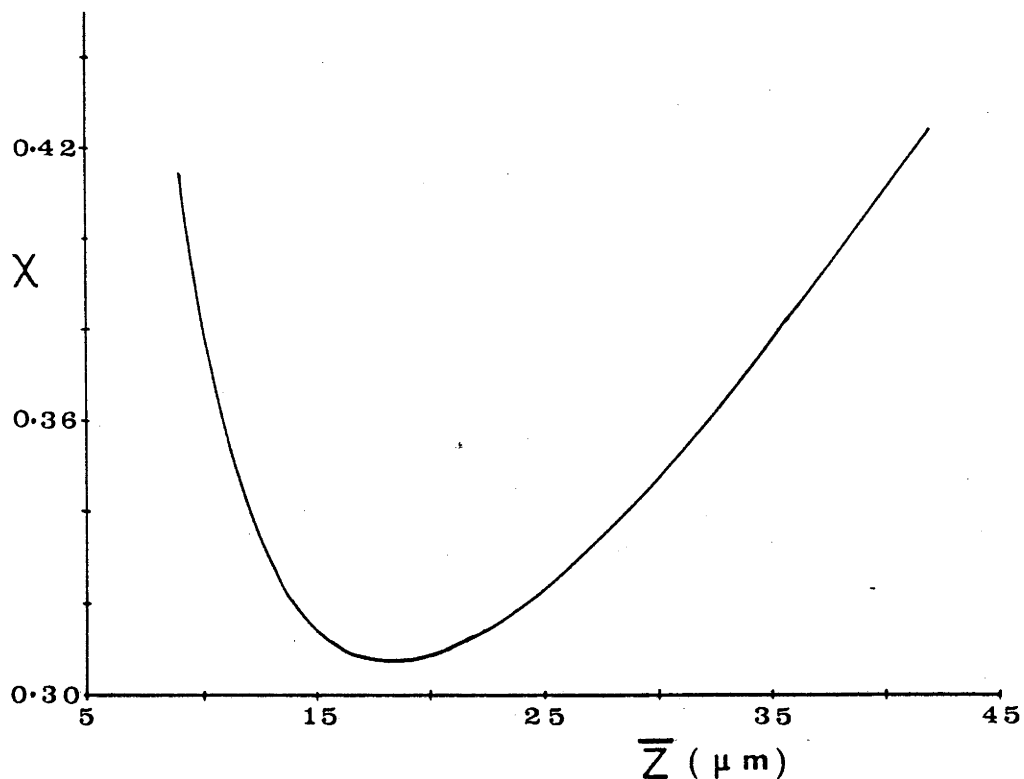


Figure 11

2. SMALL CONE-ANGLE ASSUMPTION

$$[\theta_o \sqrt{(1 + z_1^2/2\Delta)}]_{\max} = [\theta_o \sqrt{(1 + R_e^2/2\Delta)}] = 0.40$$

Note that $\sin(0.40) = 0.389$ and $\cos(0.40) = 0.921$ and hence the approximate coordinate transformation $R \approx z$ and $\theta \approx r/z$ (see Fig. 1) will be valid, with small error.

3. PARAXIAL RAY ASSUMPTION

The paraxial ray assumption (section 8.5) can be rewritten in terms of $\bar{z}$ - the midpoint of the section:

$$x(\bar{z}) = \frac{\theta_o \sqrt{[1 + (\bar{z} + L/2)^2/2\Delta][1 + (\bar{z} - L/2)^2/2\Delta]}}{(\bar{z} - L/2)/\sqrt{2\Delta}} \ll 1$$

Figure 11 is a plot of $\chi(\bar{z})$ with parameters determined for our model. Note that the maximum value of χ , within the range of $\bar{z}$ of interest, is 0.44, and that $\sin(0.44) = 0.426$, and $\tan(0.44) = 0.471$. Using the paraxial form of Snell's law is valid, with small error.

8.7.2 RAY TRAJECTORIES IN THE CONE MODEL

Some insight into the behavior and function of the Butterfly crystalline cone may be obtained by tracing the ray trajectories through the entire structure (although it must be understood that at the proximal end of the cone, the ray theory should, strictly, be supplemented by wave theory).

We consider the simplest case - rays launched radially into the cone. The paths are given by Eq. (10), and converting to cylindrical coordinates:

$$r = \theta_i z \cos\{(\sqrt{2\Delta}/z) - (\sqrt{2\Delta}/R_e)\} \quad (30)$$

where θ_i is the initial angular coordinate of the ray as it enters the cone, and R_e the radius of the cone endface (using cylindrical coordinates simplifies the problem of matching the ray trajectories in the cone and lens cylinder). The slope of the ray is given by

$$\begin{aligned} dr/dz = & (\theta_i \sqrt{2\Delta}/z) \sqrt{(1 + z^2/2\Delta)} \\ & \times \sin\{(\sqrt{2\Delta}/z) - (\sqrt{2\Delta}/R_e) + \tan^{-1}(z/\sqrt{2\Delta})\} \end{aligned} \quad (31)$$

From Eq. (30), the phase of the ray oscillation as the rays reach the lens cylinder ($z = z_1$) is given by

$$\phi = (\sqrt{2\Delta}/z_1) - (\sqrt{2\Delta}/R_e) = 1.34 < \pi/2 \quad (32)$$

using the parameter values for our model. Thus, a quarter oscillation has not quite taken place.

The rays subsequently enter the lens cylinder, Eq. (28), with trajectory

$$r = A \cos\{(\sqrt{2\Delta}/z_1^2) \tilde{z} + B\} \quad (33)$$

where A and B are constants (see, for example, Marcuse¹³). We have introduced here a new variable, $\tilde{z}$, defined by

$$\tilde{z} = z_1 - z$$

which measures distance into the lens cylinder. Differentiating Eq. (33) gives

$$\begin{aligned} dr/d\tilde{z} &= -dr/dz \\ &= -A(\sqrt{2\Delta}/z_1^2) \sin\{(\sqrt{2\Delta}/z_1^2) \tilde{z} + B\} \end{aligned} \quad (34)$$

Matching r and dr/dz at $\tilde{z} = 0$, for the ray trajectories in the cone and lens cylinder, gives

$$A = z_1 \theta_i \sqrt{1 + (z_1/\sqrt{2\Delta}) \sin 2\phi + (z_1^2/2\Delta) \cos^2 \phi} \quad (35a)$$

$$\tan B = (z_1/\sqrt{2\Delta}) + \tan \phi \quad (35b)$$

where ϕ is defined in Eq. (32). With the parameter values of our model, $A = (10.73) \theta_i \mu\text{m}$ and $B = 1.369$. From Eq. (33) it follows that $r = 0$ at $\tilde{z} = 1.12 \mu\text{m}$. Thus the rays are brought to an internal focus about a micron into the lens cylinder.

Figure 12 traces the rays through the entire model, and shows the internal focus inside the lens cylinder. Note that the rays do not emerge from the lens cylinder parallel to the axis, as would be expected if the complete optical system (corneal lens with crystalline cone) were perfectly afocal⁷⁻⁸. Small variations in the model parameters do not appear to drastically alter this result. This is contrary to the conclusions of Nilsson, Land and Howard⁷⁻⁸, who traced rays through the crystalline cone by simulating it as a series of thin lenses (one micron apart), each with focusing power given by their experimentally measured power curve (Fig. 8). They found that the rays are focused about $8 \mu\text{m}$ from the proximal end, and emerged as a parallel beam. However, their ray tracing technique is highly questionable, particularly for a graded medium as strongly tapered as the Butterfly cone.

8.8 CONCLUSION

We have described the physical properties of a new optical element which has been suggested as a model for cone structures in eyes. This involves the application of a new optical principle in vision -

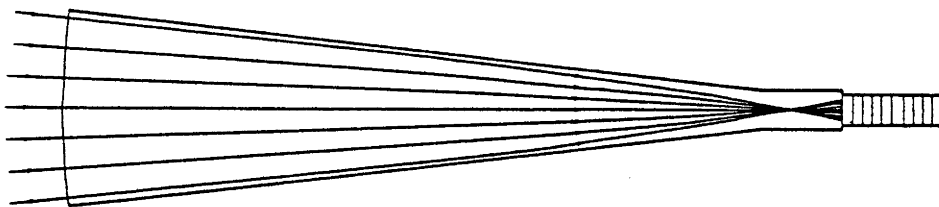


Fig. 12 Radially incident rays traced through the entire Butterfly cone structure.

lossless graded-index tapers - first introduced by Marcatili (1985) in the optical electronics field. The cone guides and concentrates light without loss, and appears particularly well suited to vision applications. We showed that the model explains the recent (unusual) measurements on the crystalline cone of a Butterfly compound eye, and leads to a new understanding of its purpose. It is interesting to note that the Butterfly eye appears to provide us with the very first concrete example of a strictly adiabatic (lossless) taper - the manufacture of these devices has not, as yet, been reported in the literature.

The full analysis of the Butterfly eye is, as yet, incomplete (for example, the angular sensitivity was only mentioned) and requires details of the photoreceptor excitation. This awaits the completion of a suitable wave theory (the special case $n_s = n_o$ is treated in chapter 3).

Finally, we could speculate about the application of our model to other eyes. The crystalline cone of *Limulus*¹⁴ (the king crab) is a tapered graded cone which conceivably could fit our model, and mention has already been made of the cone receptors in the human retina⁹. Graded cone structures appear frequently in compound eyes, but details are generally unknown.

8.9 REFERENCES

1. M.F. Land, "Compound Eyes: Old and New Optical Mechanisms," *Nature* 287, 681 (1980).
2. S. Caveney and P. McIntyre, "Design of Graded-Index Lenses in the Superposition Eyes of Scarab Beetles," *Phil. Trans. R. Soc. Lond.* B294, 589 (1981).
3. A.W. Snyder and R. Menzel (eds.), "Photoreceptor Optics," (Springer-Verlag, Berlin, 1975).
4. M.F. Land, "Superposition Images are Formed by Reflection in the Eyes of some Oceanic Decapod Crustacea," *Nature* 263, 764 (1976).
5. R. Winston, "The Visual Receptor as a Light Collector," in *Vertebrate Photoreceptor Optics*, J.M. Enoch and F.L. Tobey (eds.), (Springer-Verlag, Berlin, 1981).
6. C. Pask and D. Bertilone, "New Model for Analysing Cone Elements in Visual Systems," *Nature* (submitted).
7. D.-E. Nilsson, M.F. Land and J. Howard, "Afocal Apposition Optics in Butterfly Eyes," *Nature* 312, 561 (1984).
8. D.-E. Nilsson, M.F. Land and J. Howard, "Optical Characteristics of the Butterfly Eye," *J. Comp. Physiol.* (in press).

9. A. Sharma and I.C. Goyal, "Tapered Parabolic-Index Fiber Waveguide," Appl. Opt. 18, 1482 (1979).
10. A.W. Snyder and J.D. Love, "Optical Waveguide Theory," (Chapman and Hall, London, 1983).
11. E.A.J. Marcatili, "Dielectric Tapers with Curved Axes and No Loss," IEEE J. Quantum Electron. QE-21, 307 (1985).
12. C. Gomez-Reino and E. Larrea, "Imaging and Transforming Transmission through a Medium with Nonrotation - Symmetric Gradient Index," Appl. Opt. 22, 387 (1983).
13. D. Marcuse, "Light Transmission Optics," (Van Nostrand Reinhold, New York, 1973).
14. M.F. Land, "The Optical Mechanism of the Eye of Limulus," Nature 280, 396 (1979).