FABRY-PEROT AND WHISPERING GALLERY MODES

IN REALISTIC RESONATOR MODELS

by

DAVID H. FOSTER

A DISSERTATION

"Fabry-Perot and Whispering Gallery Modes in Realistic Resonator Models," a dissertation prepared by David H. Foster in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Physics. This dissertation has been approved and accepted by:

Dr. Hailin Wang, Chair of the Examining Committee

3/8/2006

Date

Committee in Charge: Dr. Hailin Wang, Chair
Dr. Jens Nöckel
Dr. Stephen Hsu
Dr. Richard Taylor
Dr. Michael Kellman

Accepted by:

Dean of the Graduate School.

An Abstract of the Dissertation of

David H. Foster   For the degree of   Doctor of Philosophy

in the Department of Physics   to be taken   March 2006

Title FABRY-PEROT AND WHISPERING GALLERY MODES

IN REALISTIC RESONATOR MODELS

Approved: _____

Dr. Hailin Wang, Chair

We investigate models describing two classes of microresonators: those having the shape of a dome, and those having an oval (deformed circle or sphere) shape. We examine the effects of dielectric interfaces in these structures.

For the dome cavity, we derive efficient numerical methods for finding exact electromagnetic resonances. In the dome consisting of a concave conductor and a planar, dielectric Bragg mirror, we discover a phenomenon which we call paraxial mode mixing (PMM) or classical spin-orbit coupling. PMM is the sensitive selection of the true electromagnetic modes. The true modes are generally mixtures of pairs of vectorial Laguerre-Gauss modes. While each member of an LG pair possesses definite orbital angular momentum and spin (polarization), the mixed modes do not, and exhibit rich, non-uniform polarization patterns. The mixing is governed by an orthogonal transformation specified by the mixing angle (MA). The differences in reflection phases of

a Bragg mirror at electric s and p polarization can be characterized in the paraxial regime by a wavelength-dependent quantity $\epsilon_s - \epsilon_p$. The MA is primarily determined by this quantity and varies with an apparent arctangent dependence, concomitant with an anticrossing of the maximally mixed modes. The MA is zero order in quantities that are small in the paraxial limit, suggesting an effective two-state degenerate perturbation theory. No known effective Hamiltonian and/or electromagnetic perturbation theory exists for this singular, vectorial, mixed boundary problem. We develop a preliminary formulation which partially reproduces the quantitative mixing behavior. Observation of PMM will require both small cavities and highly reflective mirrors. Uses include optical tweezers and classical and quantum information.

For oval dielectric resonators, we develop reduced models for describing whispering gallery modes by utilizing sequential tunneling, the Goos-Hänchen (GH) effect, and the generalized Born-Oppenheimer (adiabatic) approximation (BOA). While the GH effect is found to be incompatible with sequential tunneling, the BOA method is found to be a useful connection between ray optics and the exact wave solution.

The GH effect is also shown to nicely explain a new class of stable V-shaped dome cavity modes.

This dissertation includes my co-authored materials.

CURRICULUM VITAE

NAME OF AUTHOR: David H. Foster

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon

Gonzaga University

DEGREES AWARDED:

Doctor of Philosophy in Physics, 2006, University of Oregon

Master of Science in Physics, 1999, University of Oregon

Bachelor of Science in Physics, 1998, Gonzaga University

Bachelor of Science in "Math/Computer Science", 1998, Gonzaga University

AREAS OF SPECIAL INTEREST:

classical mechanics and chaos, semiclassical physics, adiabatic approximations

electromagnetism, quantum optics and information, computational physics

PROFESSIONAL EXPERIENCE:

Research Assistant, Department of Physics, University of Oregon, Eugene, 1999-2006

Teaching Assistant, Department of Physics, University of Oregon, Eugene, 1998-1999

Optical Technician, New Light Industries, Spokane, WA, 1996-1998

Laser Technician, Chemistry Department, Gonzaga University, Spokane, WA, 1997

Geochemistry Lab Intern, Pacific Northwest National Laboratory, Richland, WA, 1993-1994

GRANTS, AWARDS AND HONORS:

Summa Cum Laude, Gonzaga University 1998,

Murdock Grant, 1997

Sophmore Academic Excellence Award, 1995

Chevron Merit Scholarship and various Gonzaga University Merit Scholarships, 1994

Tandy Technology Award/Scholarship, 1994

Associated Western Universities Student Fellow, 1993-1994

National Merit Scholarship, 1993

1st Place (team) regional DOE Science Bowl competition 1994 and 1992; National placements (team) of 5th (1992) and 10th (1994) 1st place (individual and team) regional MathCounts competition, 1989

PUBLICATIONS:

David H. Foster and Jens U. Nöckel, "Bragg-induced orbital angular-momentum mixing in paraxial high-finesse cavities", Optics Letters 29, 2788 (2004).

David H. Foster and Jens U. Nöckel, "Methods for 3-D vector microcavity problems involving a planar dielectric mirror", Optics Communications **234**, 351-383 (2004).

David H. Foster and Jens U. Nöckel, "Spatial and polarization structure in micro-domes: effects of a Bragg mirror", in *Laser Resonators and Beam Control VII*, edited by Alexis V. Kudryashov and Alan H. Paxton, Proceedings of SPIE **5333**, 195-203 (2004).

Scott Lacey, Hailin Wang, David H. Foster and Jens U. Nöckel, "Directional tunnel escape from nearly spherical optical resonators", Physical Review Letters **91**, 033902 (2003).

ACKNOWLEDGMENTS

When I was six years old, my father gave me a National Geographic book titled "Our Universe". He read to me and I read to myself, and I began to tell every longsuffering adult within earshot about the solar system: distances, sizes, lengths of days, length of years, temperatures, seasons, atmospheres, names of moons, etc.... Later my father gave me math workbooks designed for several grade levels ahead of mine. I believe I was twelve or thirteen when I obtained the ultimate conceptual physics book, "Thinking Physics" by Lewis C. Epstein. So you see, this dissertation is not entirely my own fault; Dad must share some of the blame for my penchant for math and science.

There are many people who have been a help to me during my time at Oregon. I wish to say thank you to them, including many who are not explicitly named here. I want to thank my advisor, Prof. Jens Nöckel, for putting up with me, and for continual doses of brilliant insight and great help. Although the work presented in this dissertation is my own, Jens played an integral advisory role. Working with Jens has been a wonderful experience. I was hoping that his ability to indefinitly remember the concept at the heart of every physics phenomenon would rub off on me, but I have not caught this yet. I want to thank Prof. Michael Raymer, who has been a truly great help to me, both when I worked in his lab and afterward. I have greatly enjoyed our scientific and technical discussions, including the reading course on the fundamentals of quantum optics. My investigation with the microdome cavity was initially meant as a modeling tool for a CQED experiment in Raymer's lab, for which I had performed design and construction, and as a personal exercise in electromagnetic modeling as I began work with Jens. The realization that there was something left to be written about paraxial modes led to the first and most important portion of this dissertation. I hope that I continue to have opportunity to collaborate with Jens and Mike, as well as Prof. Hailin Wang, whose experimental use of whispering gallery modes overlaps with several of my investigations. Hopefully the knowledge that I have gained can be of use to them and others.

I also owe a great deal to my most extraordinary classroom teachers, Profs. Czonka, Toner, and Hsu, as well as to the teachers who were not so outstanding, but were knowledgable and helpful nonetheless. I have greatly benefited from discussions with Martina Hentschel, Avi Niv, Scott Lacey, Hailin Wang, Michael Kellman, Phillip Marston, Justin Hannigan, Brian Smith and others. I want to thank Becky and Ja-

viii

nine in the Oregon Center for Optics office for helping me travel around, print this work, and eat free food. I have found the Department of Physics, and the University of Oregon as a whole, to be wonderful academic institutions. Finally, thank you to the National Science Foundation for providing Jens and me with the funding for this research (CAREER award no. 0239332).

Of most importance, more than my research, are Brandy and Alexandra. Brandy, I love you, thank you for loving me. Alexa, I love you, but please take the ball of fuzz out of your mouth.

Table of Contents

APPENDIX

List of Figures

List of Tables

Chapter 1

INTRODUCTION

In this dissertation, we investigate numerical models describing two classes of microresonators: those having the shape of a dome, and those having an oval (deformed circle or sphere) shape. For both of these structures, we develop and analyze numerical methods of solution which include effects caused by the interactions of light with dielectric interfaces. The simplest models of these structures replace the dielectric interfaces with conductors, and loosely speaking, we are interested in effects that these models fail to produce. In the case of the dome cavity, we construct a method for the exact electromagnetic solution that is considerably less calculationally intensive than standard techniques. In the case of oval resonators, where efficient exact solution methods have been well developed, we concentrate on reduced models which can approximate the exact solutions while allowing simpler physical understanding.

Our work is connected with theoretical fields of intrinsic interest: resonances and quasinormal modes, adiabatic approximations, nonlinear dynamics and chaos, semiclassical quantization and short wave asymptotics, and the orbital angular momentum of light beams. These theoretical topics will be discussed in their relation to our microresonators. With the exception of beam orbital angular momentum, each of these topics has something to do with our analysis of both of the microresonators.

Chapters 2-4 and Sec. 6.2 discuss the dome cavity; Chapters 5-7 primarily discuss our work in oval resonators. The introductions to each microresonator problem are given at the beginnings of Chapters 2 and 5. The work on the dome cavity is the most completely developed work, while the oval resonator portion expresses current progress in an ongoing project.

Chapter 2 presents a set of novel numerical methods which we have developed to

explore the dome cavity and to see how realistic dielectric mirrors may affect modes in a more general class of cavities.

In Chapter 3 we investigate, using analytic arguments and the numerical methods developed in Chapter 2, the phenomenon of spin-orbit coupling (mixing) of paraxial modes in the dome cavity. Our work in this area is, we believe, the first investigation, theoretical or experimental, that explores how the immediate (perturbatively zero-order) partition of mode families (families which are degenerate in the paraxial approximation) is chosen in the presence of a dielectric mirror. Here we also discuss the predicted requirements for observing the mixing as well as possible applications. The investigation presented in Chapter 3 is perhaps the most important of those described in this work. The conclusion of Chapter 3 is the central, definite conclusion of the research regarding the dome cavity.

In Chapter 4 we discuss the mode mixing results of Chapter 3 in light of several reduced models. The goals of the reduced models are 1) to reproduce the mixing angle behavior found in the exact calculations, and 2) to explain this behavior in physical or familiar terms. A satisfactory reduced model which meets both goals has yet to be found, and the purpose of Chapter 4 is to illuminate several approaches toward a more analytic explanation of spin-orbit coupling and to show the current state of each approach. The first reduced model simply uses the reduced basis consisting of the two modes which are allowed to mix. Under certain conditions, this simple reduced model reproduces the exact numerical results, but does not lend a familiar interpretation to the result. The second reduced model forces the mixing problem into the familiar context of degenerate perturbation theory via an ad hoc procedure. The result of this theory surprisingly contains an essential qualitative aspect of the mixing behavior, but the theory is manifestly lacking an unknown extension which would simultaneously validate the procedure and provide a quantitative prediction. Chapter 4 concludes with an outlook discussing other models that might be used to explain paraxial mode mixing.

Chapter 5 begins the second part of the dissertation, which is primarily focused on whispering gallery modes in dielectric oval resonators. The sequential tunneling model is introduced and the Goos-Hänchen effect is explained and calculated, with the reader being referred to Appendix E for the most sophisticated calculation. Chapter 5 also gives the calculation of a different non-specular effect, which we call the Fresnel kick.

The first section of Chapter 6 investigates whether the Goos-Hänchen (GH) effect

can be added to the sequential tunneling model as an improvement to the classical dynamics. It is found that a strong inequality necessary for a reasonably clean definition of the GH effect is violated. Thus for this section we have the negative result that there is not a satisfactory method of incorporating the GH effect into the sequential tunneling model. The last section of Chapter 6 demonstrates that the GH effect has a large effect (the creation of new modes) in the non-paraxial dome cavity. In this application, the calculation of the GH effect is simple, and its predictions are accurate.

Chapter 7 investigates another reduction of exact wave calculations for whispering gallery modes in open resonators, via a generalized Born-Oppenheimer approximation. We develop a procedure called the BO-WKB method for whispering gallery modes. We analyze the requirements of the approximations involved and demonstrate the use of the BO-WKB method on modes of the dielectric ellipse. The method is shown to accurately predict the real part of the wavenumber k, and to less accurately give the imaginary part. The BO-WKB procedure is more clean and controlled than sequential tunneling, and serves as a bridge between the latter and the exact wave calculations.

Chapter 8 takes a brief second look at each topic. Chapter 8 also discusses more tenuous links to adiabatic approximations.

The next section of this chapter introduces some basic vocabulary. The rest of this chapter is devoted to the topic of resonances, which is central to this work and is important in most branches of physics.

1.1 Terminology

This section gives a *brief* glossary of terms. Many of these terms are found in the chapter and section titles, and this section is partially meant to help the casual reader navigate the table of contents. It is also meant to narrow down the definition of multi-use terms to the parlance used in this dissertation.

Bouncing Ball Mode: Waves that trace out a relatively simple and obvious closed orbit, and that in principle can travel through the central region of the billiard, are known as bouncing ball modes. A Gaussian mode in a standard Fabry-Perot laser cavity is a bouncing ball mode (the simplest possible), as is the bow-tie shaped orbit studied in Ref. [1]. Bouncing ball modes are defined in contrast to

whispering gallery modes and "chaotic" modes.

Caustic: A curve to which every ray in a certain bundle of rays is tangent. The ray density grows to infinity at the caustic, and if the rays represent a wave, the wave amplitude will be large near the caustic.

Cavity: "Cavity" means a cavity resonator. *Loosely speaking a cavity resonator is a structure that is sufficiently enclosed by an outer surface (which may technically be open) so that field resonance states (modes) exist.* A trumpet is a cavity, as is a bubble in a hunk of metal, or two concave mirrors facing each other. Sometimes the term cavity resonator is extended to include less "hole-like" optical resonators in which the inside material has a larger refractive index than the outside material, such as the oval microresonators discussed in this work. *See Sec. 1.2 for more about wave resonators.*

Fabry-Perot: A Fabry-Perot resonator is a cavity formed either by two parallel, planar mirrors or by two mirrors, one or both of which are curved, facing each other to support the two-bounce linear bouncing ball mode.

Gaussian Beams/Modes: A Gaussian beam or mode is a paraxial (wave) beam or mode which decays in directions transverse to the axis with a Gaussian envelope, $e^{-a\rho^2}$. See paraxial.

Goos-Hänchen Effect: An effect due to interference in which the center of a light beam is shifted laterally upon reflection from a dielectric interface. There are quantum mechanical and acoustic versions of the GH effect as well.

Mixing: Mixing here refers to the perturbation-induced *fixed* linear transformation of several unperturbed modes, whose unperturbed energies are equal, into several new, perturbed modes. In other words, mixing refers to what happens in zero-order degenerate perturbation theory. The term mixing is used in contrast to superposition, which refers to the *arbitrary* linear combination one can make of modes that either are truly degenerate or are overlapping due to resonance widths. We will actually be using the word "mixing" with a less precise meaning than that given above; our use of mixing involves a linear transformation of the *transverse mode patterns*, which depend on x and y and do not have an energy associated with them, rather than the full modes, which depend on x, y, and z

and do have a particular energy. Nevertheless, the precise definition sheds light on the mixing of transverse mode patterns.

Paraxial: Paraxial means "along the axis". A paraxial beam or mode has intensity localized near one axis (usually taken to be the z axis).

Paraxial Mode Mixing: See Spin-Orbit Coupling.

Poincaré surface of section: A phase space trajectory (trajectory = path of evolution of an inital condition) of an *autonomous* (time independent) Hamiltonian system with $2N$ degrees of freedom (N canonically conjugate (p, q) pairs) can be better visualized by creating a Poincaré surface of section (SOS). To accomplish this, one position variable, q_j, is picked for which typical trajectories repeatedly intersect the hyperplane $q_j = 0$. (In general, the intersection of the full trajectory $(p_1(t), q_1(t), \ldots, p_N(t), q_N(t))$, which is a 1D curve, intersects $q_j = 0$ at a set of 0D points.) The SOS is the $(2N - 2)$-D scatter plot of the subset of the intersection points that have $p_j > 0$; the axes of the plot are the p_i and q_i for $i \neq j$. For $N = 2$, the SOS is 2D, and is indispensible for the visualization of these dynamical systems.

Semiclassical: "Semiclassical" (or "quasiclassical") refers to theory based on the results of path integral quantum mechanics in the limit of small (but non-zero) $\hbar$. In semiclassical theory one sums over all *classical* paths using the Van Vleck propagator [2] with the additional feature that at each crossing of a caustic a multiple of $\pi/2$ is added to the phase. For discussions of semiclassical theory, see Chapter 7 of Ref. [3] and also Ref. [4]. The term semiclassical is sometimes used in a broader sense to refer to various treatments in which the wavelength is taken to be small. The meaning of semiclassical here should not to be confused with its common usage in quantum optics, where it refers to a treatment in which electrons are quantized and the electromagnetic field is classical.

Sequential Tunnelling Model: In a sequential tunnelling model, a mode in a dielectric resonator is modeled by a classical trajectory which loses amplitude at each reflection from a dielectric interface.

Spin-Orbit Coupling: In this work "spin-orbit coupling" or "paraxial mode mixing" refers to a specific kind of mode mixing between two modes (or two degenerate pairs of modes) which have different spin and different orbital angular

momentum quantum numbers. The resulting modes of course do not have a definite spin or orbital angular momentum (although the total angular momentum is well defined). Our results indicate that spin-orbit coupling can be caused by the effective birefringence of distributed Bragg reflectors.

Stable/Unstable Orbit/Trajectory/Mode: An unstable trajectory is a periodic orbit with the property that a slight perturbation of starting conditions generally results in the trajectory moving exponentially fast away from its original orbit. In other words, *chaotic motion* exists in the vicinity of unstable orbits. On the other hand, perturbing a stable orbits results in an orbit that stays close to the original orbit, generally spiraling around it in phase space — a *regular motion*. A stable or unstable mode is a wave mode that is associated[1] with periodic orbits that are stable or unstable. Linearizing the effect, in phase space, of small perturbations of periodic orbits results in a *monodromy matrix* which has the property of being *symplectic*. The monodromy matrix has an even order and its eigenvalues come in reciprocal pairs: λ, $1/\lambda$. If all of the eigenvalues of the monodromy matrix lie on the unit circle, the orbit is stable. Otherwise, at least one eigenvalue will have a magnitude greater than 1 and the orbit is unstable. Each eigenvalue λ_j of the monodromy matrix is related to a Lyapunov exponent α_j of the periodic orbit (of period T) by $|\lambda_j| = e^{\alpha_j T}$.

Quantum Chaos (Quantum Chaology): Quantum Chaos, or, as Michael Berry calls it more correctly, quantum chaology, refers to the quantum mechanical behavior of systems which possess a classical counterpart that is chaotic.

Resonance, Resonance State: See Sec. 1.2.

Whispering Gallery Mode: A whispering gallery (WG) mode is any wave mode that travels along the perimeter or edge of a cavity resonator. In 2D the term "WG mode" refers either to any wave that circulates in one sense only about the cavity center, or to any standing wave with counterpropagating parts that themselves circulate about the center in only one sense. Whispering gallery modes are defined in contrast to bouncing ball modes.

---

[1] The association of modes with periodic orbits is not always possible and is never strictly correct. (Even in integrable systems the mode is actually related to a particular torus in phase space, which is a quasiperiodic orbit. Within the torus there exists a special stable periodic orbit, which we say the mode is associated with). Nevertheless this loose association is often very useful.

1.2 Resonances in Wave Mechanics

In wave mechanics (optics, acoustics, quantum mechanics), resonances are peaks in the intensity of an outgoing wave with respect to the frequency of the incoming wave, where outgoing and ingoing are defined in respect to some localized system which is called the resonator or scatterer. The intensity peak for the outgoing wave is caused by the constructive interference of waves that have spent different periods of time "inside" the scatterer/resonator. Inside the resonator, in addition to traveling waves that depend strongly upon the structure of the incoming wave, there will exist a characteristic wave, built by constructive interference, the magnitude of which peaks at the resonant frequency, and the spatial structure of which changes little as the incoming wave is structurally modified. We will use the word "resonance" both to refer to the peak in output verses input frequency and to the frequency or wavenumber at which these peaks occur. We will use "resonant state" or "mode" to refer to the characteristic internal wave that exists inside the resonator at and around the resonant frequency; other synonyms include quasimode, which stresses the openness/loss aspect, and quasinormal mode, which refers to the existance of a theory [5, 6] for the generalized orthogonality and completeness of these open modes. Because of the robustness against changes in excitation, the resonance state is of great interest. (One cannot always know or control how a system will be excited, but a sufficiently isolated mode depends predominately on the material configuration of the resonator itself, which most likely does not change in unknown ways.)

For optical resonators we generally speak about the wavenumber k rather than the frequency. In this work, the real part of k will refer to the wavenumber in free space, so that $\operatorname{Re} k = 2\pi/\lambda_0$. The positive quantity $-2\operatorname{Im} k$ is equal to the FWHM (full width at half maximum) of the resonance peak on a power vs. frequency plot. This is shown in the next section in the context of scattering theory. For an introduction to resonances including temporal properties such as the Wigner delay time, see [7].

1.2.1 The Scattering Matrices and the Resonance Problem in 2D

As mentioned in the previous section, we can interpret a mode as being a wave that constructively interferes with itself after internal scattering and propagation[2]; the mode is the eigenfunction, with eigenvalue 1, of some *internal* scattering matrix, so that the internal wave, represented by vector a, reconstructs itself:

$$S_i a = a \qquad (1.1)$$

In order to cause the eigenvalue to be 1, k must take on its isolated, complex resonance values. At these resonances, the more well known *external* scattering matrix has a singularity in the sense that the amplitude of the outgoing wave can be made arbitrarily larger than the amplitude of a general incoming wave by having k approach a complex resonance. The internal and external scattering matrices arise straightforwardly as discussed below.

If one is seeking the modes inside a closed conducting resonator, the question that one asks is "What nonzero field (assumed to satisfy the appropriate homogeneous wave equation) satisfies the boundary conditions?" If the resonator is open (in the sense of a topologically open surface, and/or in the sense of the boundaries being dielectric interfaces or other transmissive surfaces), then the interior field is related to the incoming and outgoing waves in the infinite surrounding medium. In this case the resonance problem is formulated as the question, "What nonzero field (interior and exterior) satisfies all of the boundary equations and has no *incoming* wave at infinity?" (The property of there being no incoming wave in the far field is sometimes called the Sommerfeld or radiation boundary condition.)

Here we will demonstrate the scattering formalism, solving the resonance problem for a scalar field in a 2D oval dielectric, which corresponds physically to the problem of finding the electromagnetic modes, of a infinite dielectric cylinder of oval cross section, that are polarized so that E is parallel to the cylinder axis (see Fig. 1.1). The oval boundary δG, can be parametrized by the single coordinate ϕ. The Helmholtz

---

[2]While it is rarely possible to *explicitly* show, as can be done for the Fabry-Perot cavity which consists simply of two parallel planar mirrors, how a mode constructively interferes with itself inside the resonator, this interpretation is always valid.

Figure 1.1: Equivalence of 2D and 3D dielectric resonator problems.

$$\vec{E} = \hat{z}\, E_z(\rho, \phi) \qquad\qquad \psi(r, \phi) \equiv E_z(\rho, \phi)$$
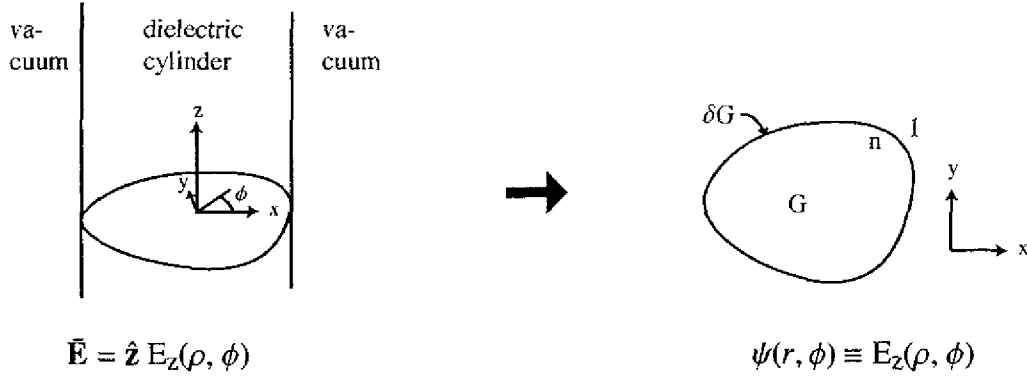
equations for ψ_i and ψ_e, the internal and external fields, are

$$\nabla^2\psi_i + n^2k^2\psi_i = 0,$$
$$\nabla^2\psi_e + k^2\psi_e = 0. \tag{1.2}$$

The complete expansion for the internal field is

$$\psi_i = \sum_{m=-\infty}^{\infty} a_m J_m(knr)e^{im\phi}, \tag{1.3}$$

where n is the index of refraction of the cylinder (which is surrounded by vacuum). This well established decomposition would include the set of Neumann functions (Bessel functions of the second kind), $Y_m(knr)$, if the origin did not lie inside G. The Y_m functions have a pole of order m (or a logarithmic singularity for $m = 0$) at the origin which precludes its use in expanding the field where there are no sources. The complete expansion for the external field, before applying the Sommerfeld boundary condition, is

$$\psi_e = \sum_{m} b_m H_m^{(1)}(kr)e^{im\phi} + c_m H_m^{(2)}(kr)e^{im\phi}. \tag{1.4}$$

Here $H^{(1)}$ and $H^{(2)}$ are the Hankel functions given by $H_m^{(3/2\mp1/2)}(x) \equiv J_m(x) \pm iY_m(x)$. In the limit $r \to \infty$, $H_m^{(1)}(kr)$ becomes proportional to e^{ikr} and is an outgoing wave, while $H_m^{(2)}$ is an incoming wave. Maxwell's equations at the cylinder surface force E_z and H to be continuous (assuming the dielectric has $\mu = \mu_0$). The continuities of E_z

and H_ϕ at the surface of the oval cylinder correspond respectively to the continuities of ψ and $\mathrm{d}\psi/\mathrm{d}r$ on δG, which respectively yield,[3]

$$\sum_m a_m J_m(knr_G(\phi))e^{im\phi} = \sum_m b_m H_m^{(1)}(kr_G(\phi))e^{im\phi} + c_m H_m^{(2)}(kr_G(\phi))e^{im\phi},$$

$$n\sum_m a_m J_m'(knr_G(\phi))e^{im\phi} = \sum_m b_m H_m^{(1)\prime}(kr_G(\phi))e^{im\phi} + c_m H_m^{(2)\prime}(kr_G(\phi))e^{im\phi}, \quad (1.5)$$

where the primes denote the derivative with respect to the entire argument (kr or knr), and $r_G(\phi)$ specifies δG. Multiplying these equations by $e^{-im'\phi}$ for each integer m' and integrating over ϕ yields the matrix equations

$$[J]a = [H^+]b + [H^-]c,$$

$$n[J']a = [H^{+\prime}]b + [H^{-\prime}]c, \quad (1.6)$$

where

$$[H^\pm]_{mm'} = \int \mathrm{d}\phi\, H_m^{(3/2\mp 1/2)}(kr_G(\phi))e^{i(m-m')\phi},$$

$$[H^{\pm\prime}]_{mm'} = \int \mathrm{d}\phi\, H_m^{(3/2\mp 1/2)\prime}(kr_G(\phi))e^{i(m-m')\phi} \quad (1.7)$$

and $[J]$ and $[J']$ are defined similarly.

If we eliminate a from (1.6) we obtain the relationship between b and c:

$$\left([J]^{-1}[H^+] - \frac{1}{n}[J']^{-1}[H^{+\prime}]\right)b = \left(\frac{1}{n}[J']^{-1}[H^{-\prime}] - [J]^{-1}[H^-]\right)c. \quad (1.8)$$

For the resonance problem, we want the outgoing wave b to be nonzero while the incoming wave c is 0. This is allowed when

$$\det\left([J]^{-1}[H^+] - \frac{1}{n}[J']^{-1}[H^{+\prime}]\right) = 0. \quad (1.9)$$

This equation can be called the *resonance condition*. Since (1.9) is an equation that sets a complex function of k equal to 0, there will generally exist solutions at discrete,

---

[3]The second correspondence comes from the Maxwell equation involving $\nabla \times E$. Note that $\mathrm{d}\psi/\mathrm{d}r$ is generally not the normal derivative. The ϕ derivative of ψ is also continuous on δG, but this constraint is not independent of the other two, as can be inferred from the fact that the tangential derivative of ψ along δG must be continuous if ψ is continuous.

complex values of k. The external scattering matrix is defined by $b = S_e c$, yielding

$$S_e = \left([J]^{-1}[H^+] - \frac{1}{n}[J']^{-1}[H^{+\prime}]\right)^{-1}\left(\frac{1}{n}[J']^{-1}[H^{-\prime}] - [J]^{-1}[H^-]\right). \tag{1.10}$$

Thus the scattering matrix has singularities (first order poles, generically) at the resonant values of k.

The internal scattering matrix is obtained by setting $c = 0$ in Eq. (1.6), eliminating b, and comparing with (1.1):

$$\begin{aligned} S_i &= \frac{1}{n}[J']^{-1}[H^{+\prime}][H^+]^{-1}[J] \\ &= \left(\frac{1}{n}[J']^{-1}[H^{+\prime}]\right)\left([J]^{-1}[H^+]\right)^{-1}. \end{aligned} \tag{1.11}$$

In order for Eq. (1.1) to have a nonzero solution a, at least one eigenvalue of S_i must be 1, which means that

$$\det\left[\left(\frac{1}{n}[J']^{-1}[H^{+\prime}]\right)\left([J]^{-1}[H^+]\right)^{-1} - I\right] = 0. \tag{1.12}$$

Multiplying this equation on the right by $\det([J]^{-1}[H^+])$ leads to Eq. (1.9). Thus the resonance condition for the internal scattering picture is the same as for the external scattering picture.

The resonance problem, as we have stated it, requires that energy be continually flowing out from the resonator while none goes into it. This is accomplished, hypothetically, by having a gain inside the cavity that exactly offsets the loss of the cavity in order to have a steady state field. Since the wavefront of the field, (or its propagating components) advances as $e^{ik \times \text{distance}}$, a negative imaginary part of k will cause the field to grow in magnitude as it travels in the resonator. Thus the interpretation of the imaginary part of the resonant values of k is that its magnitude represents the loss of the mode, and it is always the case that

$$\operatorname{Im} k \le 0, \tag{1.13}$$

with the equality holding for a perfectly closed resonator.

The relation of $\operatorname{Im} k$ to the resonance width can seen by the following argument, the latter part of which is due to [8]. Assume we start at time 0 with some amplitude for the mode in the cavity and let it decay (no hypothesized gain to offset the loss).

If the wavefront propagates at speed c/n, then the time dependence of the field at the wavefront (and at every wavefront) is of the form

$$E_{\text{front}} = E_0 e^{-|\operatorname{Im} k|(c/n)t}, \tag{1.14}$$

and the time dependence of the field at every fixed point is of the form

$$E(t) = E_0 e^{-|\operatorname{Im} k|(c/n)t} e^{-i\omega_0 t}, \tag{1.15}$$

where $\omega_0 = \operatorname{Re} k(c/n)$. Now, we wish to find the amplitude of the Fourier components $E(\omega)$ for the time-decaying oscillation given by (1.15). The Fourier decomposition of $E(t)$ is

$$E(t) = \int_{-\infty}^{\infty} E(\omega) e^{-i\omega t} \, d\omega. \tag{1.16}$$

Inverting this and using (1.15) yields

$$E(\omega) = \frac{1}{2\pi} \int_0^{\infty} E_0 e^{-|\operatorname{Im} k|(c/n)t} e^{i(\omega - \omega_0)t} \, dt. \tag{1.17}$$

Performing the simple integration and taking the complex square yields a Lorentzian line shape for the power spectrum

$$|E(\omega)|^2 \propto \frac{1}{(\omega - \omega_0)^2 + (|\operatorname{Im} k|c/n)^2}. \tag{1.18}$$

The Lorentzian $1/[(x - x_0)^2 + y^2]$ has a FWHM of $2y$, so the frequency FWHM of the resonator is $2|\operatorname{Im} k|c/n$. Multiplying by n/c gives the resonance FWHM in k-space as $-2\operatorname{Im} k$.

We note that one numerical method of solving the resonance problem is simply to attempt to satisfy boundary conditions such as (1.5) at a finite collection of points on the boundary δG. This is the method we use in practice for the exact solutions of the 2D oval microresonators in the second half of this thesis. A generalized version of the scattering matrix methods, however, is used in Variant 2 of the two-basis method discussed in Chapter 2.

1.2.2 The Quality Factor, Q

The quality factor, Q, is an important figure of merit for resonances. The universal textbook definition (cf. [8]) of Q for a monochromatically and resonantly driven system is defined in terms of energy:

$$Q \equiv \frac{2\pi \langle \text{internal (stored) energy} \rangle_{\text{cycle}}}{\text{energy loss per cycle}} \tag{1.19}$$

$$= \omega_0 \frac{\langle \text{internal (stored) energy} \rangle_{\text{cycle}}}{\text{power loss}}. \tag{1.20}$$

Following Jackson, we see from this definition that the time derivative of the physical stored energy U is given by

$$\frac{dU}{dt} = -\frac{\omega_0}{Q}U \tag{1.21}$$

with the solution

$$U(t) = U_0 e^{-(\omega_0/Q)t}. \tag{1.22}$$

The electric field then must decay as

$$E \propto e^{-(\omega_0/2Q)t}. \tag{1.23}$$

Using (1.15) we see that Q corresponds to

$$Q = -\frac{\text{Re}\, k}{2\text{Im}\, k}, \tag{1.24}$$

where k is a resonance.

The quality factor of a Fabry-Perot resonator with highly reflective mirrors that have field reflectivities r_1 and r_2 is given by [9]

$$Q \approx \frac{\text{Re}(k)L\sqrt{|r_1 r_2|}}{1 - |r_1 r_2|}, \tag{1.25}$$

where L is the effective length of the cavity.

Chapter 2

NUMERICAL METHODS FOR THE DOME CAVITY

Our primary motivation to study linear optical microcavities (two-mirror cavities with lengths of one half to hundreds of wavelengths) comes from the present and future applications of these devices. These microcavities, depicted in Fig. 2.1, can be used as mode filters for beams [10], as lasers [11], and potentially as devices which process or transform quantum information. Efforts toward achieving the latter in such cavities [12, 13, 14, 15] require *strong coupling* between a single localized electronic quantum system and a single cavity photon. In the experiments described in Refs. [12, 13, 14], the electronic system is an atom, while the experiment by Michael Raymer [15] uses an interface fluctuation quantum dot [16, 17] (an exciton in a 3D well) which is embedded in the top layer of the planar dielectric mirror. Additionally, [18] describes analytical solutions for the case of a hemiconfocal[1] parabolic dome cavity that was constructed to demonstrate quantum optical effects. Applications may extend to other cavity designs such as vertical cavity surface emitting lasers (VCSELs), for which there is current work in the calculations of the vectorial electromagnetic modes [20, 21].

Practical applications aside, the electromagnetic modes themselves are theoretically interesting. In particular, there is a relatively new understanding of beams and modes with orbital and spin angular momentum [22, 23]. These modes are discussed in the next chapter and in Appendix B. In this chapter we describe the model and numerical methods we have invented to efficiently solve for the electromagnetic modes of this class of realistic cavities. Modes over 100 wavelengths have been calculated

---

[1]See page 478 of Ref. [19].
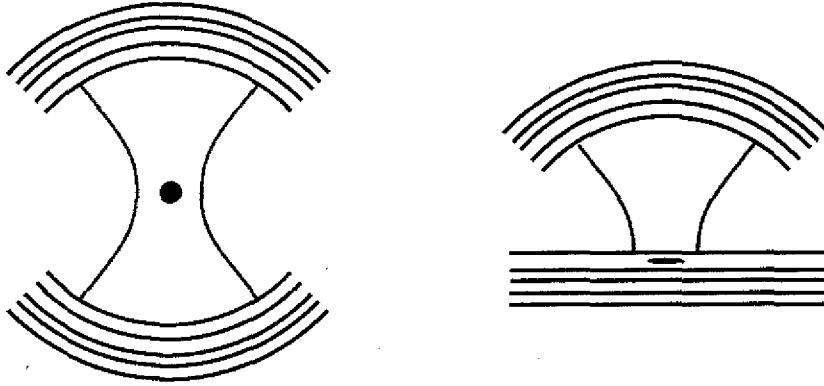
Figure 2.1: Cavities with dielectric mirrors. The envelope of the EM field is shown in light gray. The electronic quantum systems are dark gray.

with these methods. A thorough and notationally precise discussion of the model and methods is given in Ref. [24]. However, instead of repeating many pages of material in full, we somewhat shorten the discussion here, brushing over a few subtleties and occasionally using the same symbol to mean two slightly different quantities.

2.1 Introduction to the Problem

Our study of the dome cavity [24, 25, 26] has primarily been a study in electro-magnetism. While cavity resonators have been much studied using scalar fields, and the paraxial limit of this theory, Gaussian beams and modes, is a long established workhorse for the laser community [27], we focus specifically on: (a) solutions which require the vectorial nature of the electromagnetic fields and/or (b) solutions which exist in optical cavities made with dielectric mirrors, but do not exist in cavities with conducting mirrors. The latter requirement comes from practicality: at optical frequencies, the mirrors with the highest reflectivities and the lowest absorption and scattering losses are dielectric mirrors, not conducting mirrors; dielectric mirrors are ubiquitous in any field of research involving lasers.[2]

The dielectric mirrors are laminar stacks of dielectric materials, typically with alternating optical indices of refraction. Such stacks are essential 1D crystals and the principle of operation is the constructive interference of plane waves reflected from the different dielectric interfaces; hence these mirrors are also known as distributed

---

[2]At microwave frequencies, however, superconducting mirrors are available and are often used in cavity experiments.

Bragg reflectors (DBR's) or as Bragg mirrors. A planar dielectric stack of arbitrary 1D structure and infinite lateral extent yields a wonderfully simple theory of reflection and transmission of EM plane waves. A detailed explanation is given in Ref. [28]. The operation of the stack is characterized by two 2×2 transfer matrices, T_s and T_p, for s- and p-polarized plane waves. Each transfer matrix is a function of the angle of incidence, θ_k, and the wave number in free space, Re k. The matrix elements contain the plane wave reflection and transmission functions. We will primarily use the plane wave reflection functions $r_s(\theta_k)$ and $r_p(\theta_k)$, suppressing the k dependence. We will always be dealing with the monochromatic problem, where all excitations have a single angular frequency $\omega = c\mathrm{Re}\,k$.[3] The time dependence of all field quantities is $e^{-i\omega t}$ and will usually be suppressed. Our phase convention in the determination of r_p is such that, for a conducting surface, $r_p = r_s = -1$ (the other commonly used phase convention has $r_p = -r_s = +1$).

As Bragg mirrors have nonzero transmission, the optical cavities are necessarily open, or lossy. The methods described here deal with this openness correctly, with isolated complex wavenumbers k which denote both the optimal driving frequency and the resonance width. For many modes, there is also significant loss due to lateral escape from the sides of the cavity. While our model intrinsically incorporates the openness due to lateral escape in the calculation of the fields (by simply not closing the curved mirror surface, or extending its edge into the dielectric stack), this loss is not included in the calculated resonance width or quality factor, Q. Because a single set of basis vectors will be used to describe the field in the half-plane above the planar mirror, this entire half-plane is the "cavity" as far as the calculation of resonance width is concerned. We will mostly be interested in modes in which the lateral loss is negligible.

While the Bragg mirror is primarily responsible for the openness of our model system, the openness is not primarily responsible for mode pattern changes that would result from replacing the Bragg mirror with a conducting mirror. The phase shifts of plane waves reflected off of a dielectric stack can vary with incident angle, and it is this variation which can cause significant changes in the modes, even though reflectivities may be greater than 0.99. Generally speaking, the deviation of $|r_{s/p}(\theta_k)|$ from 1 is not as important as the deviation of $\arg(r_{s/p}(\theta_k))$ from, say, $\arg(r_{s/p}(0))$. In

---

[3]Strictly speaking $\omega = ck$ and will be complex for open cavities. Since we are interested in steady state applications, we will use $\omega = c\mathrm{Re}\,k$. A description of the subtleties of the time dependence is given in Ref. [24].

addition, the difference of r_s and r_p, that is, the effective *birefringent* properties of the stack, can be extremely important, as will be shown in the following chapters.

There are two general approaches to solving electromagnetic modes [29, 30]. One approach is to discretize space (and perhaps time) into finite elements and to, in some sense, solve Maxwell's equations on the grid. The other approach is to expand the EM field of each dielectric region in some complete, orthogonal set of basis functions, each of which obeys Maxwell's equations, and to then enforce the correct field continuities on all interfaces and conducting surfaces. This basis expansion approach discretizes a space of smaller dimension than does the finite element approach and this suggests it is the preferable way to proceed. The methods developed here follow the basis expansion approach with an important simplification: we will only expand the field in a single homogeneous dielectric region, rather than in all of them. The effects of the other dielectric regions, which are the layers of the stack, are correctly and efficiently handled via the transfer matrices T_s and T_p. We wish to acknowledge an excellent paper by Bava, Debernardi and Fratta [20] describing the generalization, via coupled mode theory, of the use of the plane wave and Bessel wave bases, which are discussed in this chapter, to include both gain and lateral variation of the index of refraction in order to describe Vertical Cavity Surface Emitting Lasers (VCSELs). These generalizations could be included in future work on *small* dome cavities, as they are more numerically intensive.

2.2 The Cavity Model

We can use the simple theory of dielectric stacks only for planar stacks. For this reason we use a dome-shaped cavity with the planar mirror (M1) being a dielectric stack and the curved mirror (M2) being a perfect conductor. This cavity model, combined with a basis expansion method, is, we believe, by far the least numerically intensive cavity calculation which can include the effects of a dielectric stack. A diagram of the model is shown in Fig. 2.2. The conducting surface M2 is indicated by the heavy line. The dome is cylindrically symmetric with maximum height $z = z_d$ and edge height $z = z_e$. The shape of the dome is arbitrary, but in our results the dome will be a part of an origin-centered sphere of radius $R_s = z_d$ unless otherwise specified. To force all EM fields to zero at the edge, there is tiny annular portion ($w_a \ll \lambda$) of M2 extending horizontally from the dome edge. The region surrounding the curved mirror will be referred to as layer 0. The dielectric interface between layer $n - 1$ and layer n has
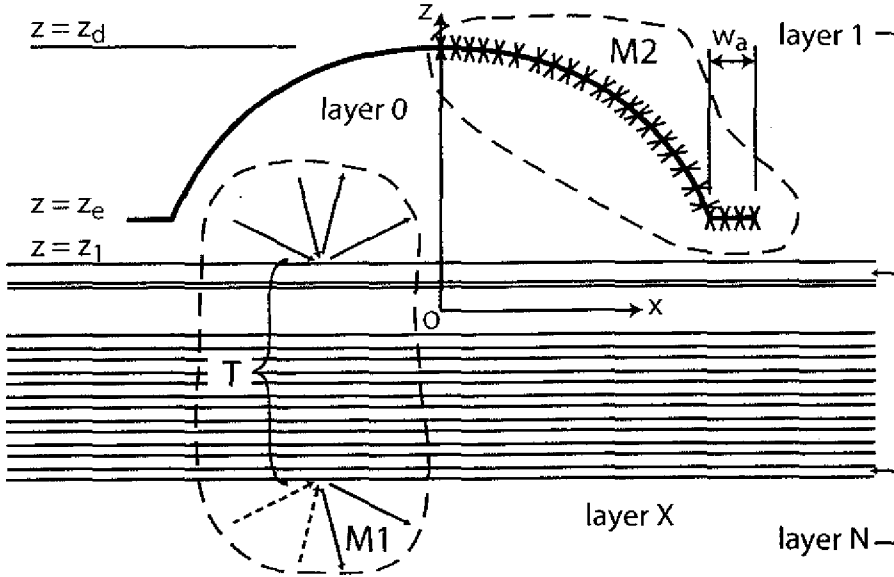
Figure 2.2: The cavity model.

height $z = z_n$. There are N layers in the stack, and the exit medium is called layer X. The depiction of the stack layers in the figure suggests a design in which the stack consists of some layers of experimental interest (perhaps containing quantum wells, quantum dots, or other structures) at the top of the stack where the field intensity is high, and a highly reflective periodic structure below. The nominal length of the cavity is $L \equiv z_d - z_1$.

At the heart of the procedure to solve for the modes is an overdetermined, complex linear system of equations, $Ay = b$. The column vector y is made up of the coefficients of eigenmodes in some basis B. The field in layer 0 is given by expansion in B using these coefficients. For a given wavenumber, k, a solution vector $y = y_{best}$ can be found so that $|A(k)\,y - b|^2$ is minimized with respect to y. We solve this linear least squares problem via a numerical library function which uses singular value decomposition [31]. Dips in the 3D graph of the residual quantity,

$$\Delta_r \equiv |A(k)\,y_{best}(k) - b|, \qquad (2.1)$$

versus complex k signify the locations of the isolated eigenvalues of k (theoretically Δ_r should become 0 at the eigenvalues). The solution vector $y_{best}(k)$ at one of these eigenvalues describes a mode. The system of equations is made up of three parts (as shown below): M1 equations, M2 equations and an arbitrary amplitude or "seed"

equation.

$$
A\,y = \begin{bmatrix} \begin{bmatrix} \begin{bmatrix} & M1 & \\ & & \end{bmatrix} \\ \begin{bmatrix} & M2 & \\ & & \end{bmatrix} \\ \begin{bmatrix} \text{s. eqn.} \end{bmatrix} \end{bmatrix} \cdot \begin{bmatrix} \\ y \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} .
$$
(2.2)

By our choosing, A usually has 2-4 times as many rows as columns.

The M1 boundary condition for a plane wave basis is expressed simply in terms of the 2×2 stack transfer matrices $T_s(\theta_k)$ and $T_p(\theta_k)$, as suggested by the M1 region (enclosed by the dashed line) in Fig. 2.2. The dashed k vectors in the figure (incoming from the bottom of the stack) represent plane waves that are given zero amplitude in order to define an eigenmode problem rather than a scattering problem. The plane waves denoted by the solid k vectors have nonzero amplitude. The nature and number of the M1 equations depends on the basis chosen. Ultimately, however, these equations enforce the stack reflection operation on plane waves:

$$
\tilde{\psi}_s(\theta_k) = r_s(\theta_k)\tilde{\psi}_s(\pi - \theta_k),
$$
$$
\tilde{\psi}_p(\theta_k) = r_p(\theta_k)\tilde{\psi}_p(\pi - \theta_k),
$$
(2.3)

where here we use the restriction $0 \leq \theta_k < \pi/2$. Here $\tilde{\psi}_{s/p}(\theta)$ is the amplitude for a plane wave of polar direction θ and polarization s or p. A plane wave is called s-polarized with respect a surface if its electric field vector is parallel to $k \times \hat{n}$ where $k \propto E \times H$ denotes the plane wave and $\hat{n}$ is normal to the surface (here $\hat{n} = \hat{z}$). A p-polarized plane wave has its magnetic field vector in the $k \times \hat{n}$ direction.

The M2 boundary condition is implemented via a "point matching method" [32] (also called a "decomposition method" [30]) as follows. A finite number of locations on the curved mirror are chosen (the "X" marks in Fig. 2.2). The M2 equations are the equations in basis B setting the appropriate fields at these locations to zero. For a problem not possessing cylindrical symmetry, these locations would be points lying on a 2D surface. The simplification due to this symmetry, however, allows these locations to be entire rings about the z axis, each ring being addressable by

a single parameter such as the ρ coordinate. There are three boundary equations at each location, corresponding to the field component equations at the surface of a conductor

$$E_\phi = 0,$$
$$E_\| = 0,$$
$$H_\perp = 0. \tag{2.4}$$

Here $E_\|$ is the component of $\boldsymbol{E}$ that is both tangent to M2 and perpendicular to $\hat{\phi}$, and $H_\perp$ is the component of the the magnetic field $\boldsymbol{H}$ that is normal to M2.

Finally, the seed equation sets some combination of basis coefficients equal to one and is the only equation with a nonzero value on the right hand side ($\boldsymbol{b}$).

We have implemented the solution for two bases for the electromagnetic field, the Bessel Wave Basis (BWB) and the Vector Multipole Basis (VMB). We will also refer to the use of the BWB as the Bessel wave method and the use of the VMB as the two-basis method (because using the VMB involves effective conversions to the BWB for parts of the calculation). Before describing these bases we will examine how the cylindrical symmetry of the boundary conditions simplifies the problem.

2.3   Basis Reduction due to Cylindrical Symmetry

Here we show that any solution of the electromagnetic field for cylindrically symmetric boundary conditions can be decomposed into a set of uncoupled 2D problems. Each of these 2D problems is associated with a particular value of m, which turns out to be the angular momentum quantum number. The solution to each 2D problem is such that E_ϕ, E_ρ, E_z, H_ϕ, H_ρ, and H_z have the form $f(\rho, z)e^{im\phi}$. Another way to say this is that the energy eigenmodes are also the eigenmodes of angular momentum. In practice, one does not even have to solve the 2D problem for many m; one simply picks the single (or few) m of interest. Fortunately, degeneracies in which two modes of different $|m|$ have the same real part of k, e.i. those which would be excited maximally by light of same frequency, are accidental.

We assume the entire EM field is a solution to Maxwell's equations and also a solution to the appropriate boundary equations describing the cylindrically symmetric dielectric interfaces and conducting surfaces. The EM field, in each homogeneous

region of space, can be expanded in an azimuthal Fourier series:

$$f(\rho, z, \phi) = \sum_m f_m(\rho, z)e^{im\phi}, \tag{2.5}$$

where f stands for any of the scalar fields $E_\phi, E_\rho, E_z, H_\phi, H_\rho, H_z$. At each dielectric interface γ separating regions a and b,

$$
\begin{aligned}
E_\phi^{(a)} &= E_\phi^{(b)} & \forall\, \boldsymbol{x} \in \gamma, \\
E_\rho^{(a)} \cos\eta - E_z^{(a)} \sin\eta &= E_\rho^{(b)} \cos\eta - E_z^{(b)} \sin\eta & \forall\, \boldsymbol{x} \in \gamma, \\
H_\rho^{(a)} \sin\eta + H_z^{(a)} \cos\eta &= H_\rho^{(b)} \sin\eta + H_z^{(b)} \cos\eta & \forall\, \boldsymbol{x} \in \gamma, \tag{2.6}
\end{aligned}
$$

where η is the polar angle of $\hat{n}(\boldsymbol{x})$, the surface normal at $\boldsymbol{x}$. At each *conducting* surface γ, these equations are modified by setting each right hand side to 0, resulting in Eq. (2.4). Substituting the form (2.5) into (2.6) and noting that we can integrate in ϕ against $e^{-im'\phi}$ yields the implication that the electromagnetic field associated with a single m will satisfy the boundary conditions (2.6) if the entire field does so.

Showing that the same implication also holds for Maxwell's equations completes the proof. Taking $\epsilon_0 = \mu_0 = c = 1$, the curl equations are

$$
\begin{aligned}
\boldsymbol{\nabla} \times \boldsymbol{E} &= i\omega \boldsymbol{H}, \\
\boldsymbol{\nabla} \times \boldsymbol{H} &= -i\omega n^2 \boldsymbol{E}. \tag{2.7}
\end{aligned}
$$

(It is not necessary to include the divergence equations here, as these follow from taking the divergence of the above equations.) The operations curl and div, *when the vector fields are given in cylindrical or spherical components*, do not contain nor generate any dependence on ϕ other than $e^{im\phi}$. For instance, the curl in cylindrical components and coordinates is

$$\boldsymbol{\nabla} \times \boldsymbol{A} = \hat{\rho}\left(\frac{1}{\rho}\frac{\partial A_z}{\partial \phi} - \frac{\partial A_\phi}{\partial z}\right) + \hat{\phi}\left(\frac{\partial A_\rho}{\partial z} - \frac{\partial A_z}{\partial \rho}\right) + \hat{z}\frac{1}{\rho}\left(\frac{\partial}{\partial \rho}(\rho A_\phi) - \frac{\partial A_\rho}{\partial \phi}\right), \tag{2.8}$$

and the harmonic ϕ-dependence, $e^{im\phi}$, is preserved. Again, integrating against $e^{-im'\phi}$ shows that the equations for different m decouple, yielding 2D equations in ρ and z. Had we chosen to use Cartesian components for $\boldsymbol{E}$ and $\boldsymbol{H}$, or any set of components that did not include E_ϕ and H_ϕ, different m values would have been coupled.

Thus $e^{im\phi}$ can be factored out of the electromagnetic field in much the same way

that $e^{-i\omega t}$ is factored out. The main complication is that a vector statement such as

$$\boldsymbol{E} = \boldsymbol{E}_m(\rho, z)e^{im\phi} \tag{2.9}$$

is ambiguous. While E_ϕ, E_ρ, and E_z are proportional to $e^{im\phi}$, E_x and E_y are not (they both contain $e^{i(m+1)\phi}$ and $e^{i(m-1)\phi}$), making the notation "$\boldsymbol{E}_m(\rho, z)$" unclear.

2.4 The Bessel Wave Basis and Method

The most natural basis to use in a problem involving a dielectric stack, which acts through (2.3), is the (vector) plane wave basis. The result of the previous section suggests that the basis can be reduced (so that the new 2D basis functions are complete in ρ and z but all have the same quantum number m). This reduction results in a new basis in which each basis function is a Bessel wave, also known as a Bessel beam or simply as a cylindrically symmetrized plane wave. This reduction/symmetrization is relatively transparent for scalar plane waves and proceeds as follows.

2.4.1 Scalar BWB

The Fourier expansion for the scalar field ψ is

$$\psi(x) = \int_0^{2\pi} d\phi_k \int_0^\pi d\theta_k \sin(\theta_k)\tilde{\psi}_k e^{i\boldsymbol{k}\cdot\boldsymbol{x}}. \tag{2.10}$$

Expanding the plane wave coefficients $\tilde{\psi}_k$ themselves as

$$\tilde{\psi}_k(\theta_k, \phi_k) = \sum_n f_n(\theta_k) e^{in\phi_k}, \tag{2.11}$$

and substituting into (2.10) yields

$$\psi = \sum_n \int_0^\pi d\theta_k \sin(\theta_k)e^{izkn\cos\theta_k}f_n(\theta_k) \int_0^{2\pi} d\phi_k \, e^{i\rho kn_0 \sin(\theta_k)\cos(\phi-\phi_k)}e^{in\phi_k}. \tag{2.12}$$

Using an integral expression for the regular Bessel function of integral order

$$J_n(x) = \frac{1}{2\pi i^n} \int_0^{2\pi} e^{ix\cos\phi}e^{in\phi} \, d\phi, \quad x > 0, \tag{2.13}$$

yields

$$\psi = 2\pi \sum_n i^n e^{in\phi} \int_0^\pi d\theta_k \sin(\theta_k) e^{izkn_0 \cos\theta_k} J_n(\rho k n_0 \sin\theta_k) f_n(\theta_k). \tag{2.14}$$

Forcing $\psi \propto e^{im\phi}$ and defining $\psi_k \equiv f_m$ yields

$$\psi = 2\pi i^m e^{im\phi} \int_0^\pi d\theta_k \sin(\theta_k) e^{izkn \cos\theta_k} J_m(\rho k n \sin\theta_k) \psi_k(\theta_k), \tag{2.15}$$

where we have dropped the "0" subscript on the refractive index. This is the expansion in scalar Bessel waves, which are given by the expression

$$2\pi i^m e^{im\phi} \exp(izkn \cos\theta_k) J_m(\rho k n \sin\theta_k), \tag{2.16}$$

with $\{\psi_k(\theta_k)\}$ being the set of coefficients.

2.4.2   Vector BWB

The discussion here is an overview of the development in Ref. [24] and only expressions for the electric field will be given (see Appendix A for the magnetic field). The electric field $\boldsymbol{E}$ can be divided into two parts: $\boldsymbol{E} = \boldsymbol{E}_s + \boldsymbol{E}_p$. That is, we divide it into the field contributions from s-polarized and p-polarized plane waves. Before symmetrization, the most compact expansion for the vector field is

$$\boldsymbol{E}_s = \int d\Omega_k \tilde{S}_k \boldsymbol{\epsilon}_{s,k} e^{i\boldsymbol{k}\cdot\boldsymbol{x}},$$
$$\boldsymbol{E}_p = \int d\Omega_k \tilde{P}_k \boldsymbol{\epsilon}_{p,k} e^{i\boldsymbol{k}\cdot\boldsymbol{x}}, \tag{2.17}$$

where $\tilde{S}_k$ and $\tilde{P}_k$ are the plane wave expansion coefficients, as $\tilde{\psi}_k$ was in the scalar case. The unit vectors are given by

$$\boldsymbol{\epsilon}_{s,k} = -\hat{\phi}_k$$
$$= \hat{\boldsymbol{x}} \sin\phi_k - \hat{\boldsymbol{y}} \cos\phi_k,$$
$$\boldsymbol{\epsilon}_{p,k} = \hat{\theta}_k \,\mathrm{sgn}(\cos\theta_k)$$
$$= \hat{\rho}_k \,\mathrm{sgn}(\cos\theta_k) \cos\theta_k - \hat{z} \,\mathrm{sgn}(\cos\theta_k) \sin\theta_k. \tag{2.18}$$

The next step is to reduce the basis, moving from $\tilde{S}_k(\theta_k, \phi_k)$ and $\tilde{P}_k(\theta_k, \phi_k)$ to $S_k(\theta_k)$ and $P_k(\theta_k)$. The result is that the electric field is given by

$$E_z(\boldsymbol{x}) = -2\pi i^m e^{im\phi} \int_0^\pi \mathrm{d}\theta_k \sin^2(\theta_k)\,\mathrm{sgn}(\cos\theta_k)e^{izkn\cos\theta_k}J_m(\rho kn\sin\theta_k)P_k,$$

$$E_\rho(\boldsymbol{x}) = \frac{1}{2}\big[S_+ e^{i\phi} + S_- e^{-i\phi} + i(P_+ e^{i\phi} - P_- e^{-i\phi})\big],$$

$$E_\phi(\boldsymbol{x}) = \frac{1}{2}\big[i(S_+ e^{i\phi} - S_- e^{-i\phi}) - P_+ e^{i\phi} - P_- e^{-i\phi}\big], \tag{2.19}$$

where

$$S_\pm = 2\pi i^m e^{i(m\mp1)\phi} \int_0^\pi \mathrm{d}\theta_k \sin(\theta_k)e^{izkn\cos\theta_k}J_{m\mp1}(\rho kn\sin\theta_k)S_k,$$

$$P_\pm = -2\pi i^m e^{i(m\mp1)\phi} \int_0^\pi \mathrm{d}\theta_k \sin(\theta_k)|\cos\theta_k|e^{izkn\cos\theta_k}J_{m\mp1}(\rho kn\sin\theta_k)P_k. \tag{2.20}$$

The quantities $S_\pm$ and $P_\pm$ are defined so that expanding the electric field into circular components gives the intuitive relation:

$$E_\pm = \frac{1}{\sqrt{2}}\big(S_\pm \pm iP_\pm\big). \tag{2.21}$$

(Here $\boldsymbol{E} = E_+\boldsymbol{\sigma}^+ + E_-\boldsymbol{\sigma}^- + E_z\hat{\boldsymbol{z}}$ where $\boldsymbol{\sigma}^\pm = (1/\sqrt{2})(\hat{\boldsymbol{x}} \pm i\hat{\boldsymbol{y}})$.)

Because each BWB function is made of plane waves with a single angle of incidence, this allows us to use the simple stack functions (2.3) directly. Thus the M1 equations in the BWB are[4]

$$S_k(\theta_k) = r_s(\theta_k)S_k(\pi - \theta_k),$$

$$P_k(\theta_k) = r_p(\theta_k)P_k(\pi - \theta_k), \tag{2.22}$$

with the restriction $0 \le \theta_k < \pi/2$. If the range $[0, \pi/2)$ is discretized into N_{θ_k} points, the number of M1 equations will be $2N_{\theta_k}$. With this choice the number of columns in A is $4N_{\theta_k}$.

The M2 equations are of course somewhat more complicated, since (2.19) must be used to enforce (2.4). Here the integral is discretized and ϕ can be set to 0. Last but not least, the seed equation often simply sets a single coefficient equal to 1 (for example, $S_k(\theta_k = \pi/16) = 1$). The system of equations is solved as described in

---

[4]This is for $z_1 = 0$. Otherwise the right hand sides must include the prefactor $\exp[-i2knz_1\cos\theta_k]$.

Sec. 2.2. After the solution is found, one may wish to calculated the field not only in layer 0, but also in the stack layers. This is accomplished by propagating the Bessel waves in layer 0 down into the stack, using the transfer matrices $T_s^{(q)}$ and $T_p^{(q)}$ for the sub-stack consisting of layers 1 through q. This is described further in Ref. [24]. One limitation of our method is that, since the stack layers often have a larger index of refraction n than layer 0, the fields in a stack layer q are incompletely expanded, always having zero amplitude for Bessel waves with polar angles within $\pm(\pi/2 - \arcsin(n_q/n_0))$ of $\pi/2$. For that matter one may ask whether the plane basis is even complete in layer 0. Strictly speaking, it is not. Berry, however, has considered the problem in 2D with a conducting, closed cavity surface [33]. In this case it is found that evanescent waves are not needed to expand the field in a finite area about the origin. Regardless of the extendibility of this proof to our model, we will be concerned only with modes where light is traveling primarily along the z axis, so the issue is not critical. This incompleteness issue arises also for the vector multipole basis.

2.5   The Vector Multipole Basis and the Two-Basis Method

The VMB has an advantage in that it is the eigenbasis of a conducting hemisphere, the "canonical" dome-shaped cavity[5]. (Of course it is most noted for being the eigenbasis of the complete sphere, and has long been used in scattering problems.) In practice, the VMB has performed better than the BWB at finding the types of modes that we look at in the next chapter (although intuitively one would think that these paraxial modes would be best expanded in Bessel waves). Using the VMB with the M1 equations (2.3) requires a effective conversion between the VMB and the BWB. Additionally, computing the fields in the stack layers after a solution y has been obtained with the VMB requires a conversion to the BWB so that the plane waves can be propagated into the stack via the transfer matrices. We use the term "two-basis method" to describe our use of the VMB because of the role of the Bessel waves in these two calculations. Fig. 2.3 represents the linear system of equations of

---

[5]Here we begin directly with the vectorial multipoles. As for Bessel waves, there is a scalar version, with the scalar multipole function being $j_l(knr)Y_{lm}(\theta, \phi)$. We have completely implemented the two-basis method for the scalar multipole basis in order to compare vector and scalar solutions. The scalar version of the Bessel wave basis/method has also been implemented.
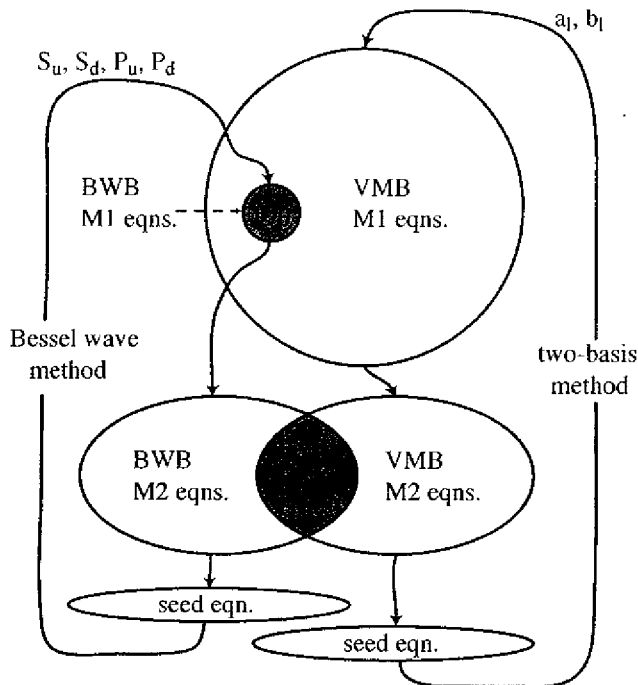
Figure 2.3: Diagram for the two methods/bases. The closed loops suggest the self-consistency or "constructive interference" of the mode solutions. Grey regions indicate intersection between BWB and VMB. Size roughly indicates the work required to get the equations. The variable coefficients are shown at the top.

the two methods and how they are related.

The complete, orthogonal vector multipole basis functions can be given as

$$M_{lm}^{(1)}(x) = -j_l(knr)\,x \times \nabla Y_{lm}(\theta, \phi),$$

$$M_{lm}^{(2)}(x) = -n_l(knr)\,x \times \nabla Y_{lm}(\theta, \phi),$$

$$N_{lm}^{(1)}(x) = \frac{1}{kn}\nabla \times M_{lm}^{(1)},$$

$$N_{lm}^{(2)}(x) = \frac{1}{kn}\nabla \times M_{lm}^{(2)}. \tag{2.23}$$

Here j_l and n_l are the spherical Bessel functions of the first and second kind and Y_{lm}

is the scalar spherical harmonic. The explicit forms of the VMB functions are

$$
\boldsymbol{M}_{lm}(\boldsymbol{x}) = \hat{\boldsymbol{\theta}} \left(\frac{im}{\sin\theta} f_l(k_n r) Y_{lm}(\theta,\phi) \right)
$$
$$
+ \hat{\boldsymbol{\phi}} \left(-f_l(k_n r) \frac{\partial}{\partial\theta} Y_{lm}(\theta,\phi) \right),
$$
$$
\boldsymbol{N}_{lm}(\boldsymbol{x}) = \hat{\boldsymbol{r}} \left(\frac{l(l+1)}{k_n r} f_l(k_{n_0} r) Y_{lm}(\theta,\phi) \right)
$$
$$
+ \hat{\boldsymbol{\theta}} \left(\frac{1}{k_{n_0} r} \frac{\partial}{\partial r} \left(r f_l(k_n r) \right) \frac{\partial}{\partial\theta} Y_{lm}(\theta,\phi) \right)
$$
$$
+ \hat{\boldsymbol{\phi}} \left(\frac{im}{k_n r \sin\theta} \frac{\partial}{\partial r} \left(r f_l(k_n r) \right) Y_{lm}(\theta,\phi) \right), \tag{2.24}
$$

where f_l denotes j_l or n_l. It is immediately seen that the EM fields, which are proportional to the above functions, have the correct components proportional to $e^{im\phi}$. The dimensional reduction due to cylindrical symmetry is accomplished by picking a single value of m instead of expanding with basis functions of many m.

We will immediately drop $\boldsymbol{M}_{lm}^{(2)}$ and $\boldsymbol{N}_{lm}^{(2)}$ from the set of basis functions we use. The reason for this is that it is necessary to convert to plane waves, and a real electromagnetic plane wave is can be complete completely expanded using $\boldsymbol{M}_{lm}^{(1)}$ and $\boldsymbol{N}_{lm}^{(1)}$. Furthermore, n_l has a singularity at the origin; thus if the origin lies in layer 0, we can eliminate these basis functions. In practice, we have not seen significant field differences between our solutions obtained with the origin inside layer 0 and for those with the origin outside layer 0 (as in Fig. 2.2).

The expansion of the electromagnetic field in layer 0 is

$$
\boldsymbol{E}(\boldsymbol{x}) = \sum_{l=l_{\min}}^{l_{\max}} \left(-a_l \boldsymbol{N}_{lm} + i b_l \boldsymbol{M}_{lm} \right),
$$
$$
\boldsymbol{H}(\boldsymbol{x}) = n_0 \sum_{l=l_{\min}}^{l_{\max}} \left(i a_l \boldsymbol{M}_{lm} + b_l \boldsymbol{N}_{lm} \right), \tag{2.25}
$$

where $l_{\min} = \max(1, |m|)$ and the expansion is truncated after $l = l_{\max}$. The a_l and b_l are complex coefficients and there are $N_l = l_{\max} - l_{\min} + 1$ of each of them. The a_l coefficients correspond to electric multipoles and the b_l coefficients correspond to magnetic multipoles [8].

The next few sections discuss the derivation of the M1 equations. The M2 equations (for both variants) come directly from the combination of (2.4), (2.25) and

(2.24), where the derivatives in (2.24) are evaluated using the commonly known formulas which related these derivatives to special functions of neighboring order.

2.5.1 Fourier Expansion using Vector Multipole Coefficients

The expansion of a scalar plane wave in terms of the *scalar* multipole basis functions, $j_l(knr)Y_{lm}(\theta, \phi)$, is given by

$$e^{ik\cdot x} = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} i^l Y_{lm}^*(\theta, \phi) j_l(kr) Y_{lm}(\theta_k, \phi_k). \qquad (2.26)$$

A discussion of the derivation of this expansion is given in Sec. 10.3 of Ref. [8]. Several numerical checks of Eq. (2.26) suggest that it is valid for complex as well as real k. The inverse of this relation, the Fourier expansion of the scalar multipole, is

$$j_l(kr) Y_{lm}(\theta, \phi) = \int d\Omega_k \left(\frac{(-i)^l}{4\pi} Y_{lm}(\theta_k, \phi_k) \right) e^{ik\cdot x}. \qquad (2.27)$$

This equation is easy to verify by inserting (2.26) into the right hand side. The use of $Y_{lm} = (-1)^m Y_{l,-m}^*$ and the orthogonality relation for the Y_{lm} leads directly to the left hand side.

We wish to use Eq. (2.27) in determining the Fourier expansion of the vector multipoles. Here we can make use of the properties of the orbital angular momentum operator $L = -i(x \times \nabla)$. The basis function M_{lm} is given by $-ij_l L Y_{lm}$. The operator L can be decomposed into components L_x, L_y, L_z where L_x and L_y can be decomposed into raising and lowering operators:

$$L_x = \frac{1}{2}(L_+ + L_-),$$
$$L_y = \frac{1}{2i}(L_+ - L_-). \qquad (2.28)$$

The action LY_{lm} can now be determined by using

$$L_+ Y_{lm} = \sqrt{(l-m)(l+m+1)} Y_{l,m+1},$$
$$L_- Y_{lm} = \sqrt{(l+m)(l-m+1)} Y_{l,m-1},$$
$$L_z Y_{lm} = m Y_{lm}. \qquad (2.29)$$

Thus

$$M_{lm} = (-ij_l)\left[\hat{x}\frac{1}{2}(d_l^m Y_{l,m+1} + d_l^{-m} Y_{l,m-1}) + \hat{y}\frac{i}{2}(-d_l^m Y_{l,m+1} + d_l^{-m} Y_{l,m-1}) + \hat{z}m Y_{lm}\right],$$

(2.30)

where

$$d_l^m = \sqrt{(l-m)(l+m+1)}.$$

(2.31)

Using (2.27) multiple times yields

$$M_{lm} = \int d\Omega_k \tilde{M}_{lm} e^{i\mathbf{k}\cdot\mathbf{x}},$$

(2.32)

with

$$\tilde{M}_{lm} = \hat{x}\frac{(-i)^l}{8\pi}(-i)e_{lm}^+ + \hat{y}\frac{(-i)^l}{8\pi}(-1)e_{lm}^- + \hat{z}\frac{(-i)^l}{4\pi}m Y_{lm}(\Omega_k),$$

(2.33)

where

$$e_{lm}^\pm \equiv d_l^m Y_{l,m+1}(\Omega_k) \pm d_l^{-m} Y_{l,m-1}(\Omega_k).$$

(2.34)

We can avoid the task of expanding N_{lm} into $j_l Y_{lm}$ terms by just using

$$N_{lm} = \frac{1}{kn}\nabla \times M_{lm} = \int d\Omega_k \frac{i}{kn}\mathbf{k} \times \tilde{M}_{lm} e^{i\mathbf{k}\cdot\mathbf{x}}.$$

(2.35)

Performing the cross product and substituting into (2.25) using (2.32, 2.33) yields the Fourier expansion of the electric field

$$E_i(\mathbf{x}) = \int d\Omega_k \tilde{E}_i(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}},$$

(2.36)

where the subscript i stands for x, y, or z and

$$\tilde{E}_x = \sum_l \frac{(-i)^l}{4\pi}\left[a_l\left(-m\sin(\theta_k)\sin(\phi_k)Y_{lm}(\Omega_k) - \frac{i}{2}\cos(\theta_k)e_{lm}^-\right) + b_l\left(\frac{1}{2}e_{lm}^+\right)\right],$$

$$\tilde{E}_y = \sum_l \frac{(-i)^l}{4\pi}\left[a_l\left(m\sin(\theta_k)\cos(\phi_k)Y_{lm}(\Omega_k) - \frac{1}{2}\cos(\theta_k)e_{lm}^+\right) + b_l\left(\frac{-i}{2}e_{lm}^-\right)\right],$$

$$\tilde{E}_z = \sum_l \frac{(-i)^l}{4\pi}\left[a_l\left(\frac{i}{2}\sin(\theta_k)\cos(\phi_k)e_{lm}^- + \frac{1}{2}\sin(\theta_k)\sin(\phi_k)e_{lm}^+\right) + b_l(mY_{lm}(\Omega_k))\right].$$

(2.37)

This Fourier expansion can now be used in the M1 equations (2.3) after deriving expressions such as $\tilde{E}(k)\cdot\epsilon_{s,k}$ and $\tilde{E}(k)\cdot\epsilon_{p,k}$. The explicit form of the resulting equations is given in [24]. There remains two different ways to obtain the equations that will be actually used in the M1 portion of the problem matrix A.

2.5.2 Variant 1

The result of the previous section are two equations (not shown) which correspond to (2.3), contain a_l and b_l, and hold for all polar plane wave angles θ_k. The simplest way to use these equations is to set $\phi_k = 0$ and pick a discrete set of θ_k values on the interval $(0, \pi/2)$, yielding $2N_{\theta_k}$ equations that will give the M1 block of A. This variant is essentially a point matching method in k space. The number of unknowns in either variant is $2(l_{\max} - l_{\min} + 1)$, and we usually pick N_{θ_k} such that there are at least as many M1 equations as unknowns.

2.5.3 Variant 2

The more elegant way to obtain the final M1 equations is to not pick discrete values of θ_k, but to integrate the preliminary M1 equations, which hold for all θ_k, against the basis $\{Y_{l'm}(\theta_k, \phi_k)\}$, which is complete for a given m. We truncate l' at $l_{\max}$, the most natural choice. This variant is analogous to the internal scattering matrix method discussed in Sec. 1.2.1 (the "S-matrix method" of Refs. [34, 35]), which integrates real space boundary conditions against the harmonic basis $\{e^{im'\phi}\}$ in two dimensions to obtain the internal scattering matrix.

It is interesting to note that, because the integration against $Y_{l'm}(\Omega_k)$ happens turns out to give 0 whenever $l' + m$ is odd, the number of M1 equations obtained this way is equal to or very close to half of the number of unknowns. It makes intuitive sense that a dielectric mirror should constrain the fields above it by half. If another dielectric mirror is placed above the present one and there are no incoming waves, the field between the mirrors, if known to have a single angular momentum quantum number m, is known to within a few free parameters (although k is not quantized). This same property of M1 providing half of the system constraints is explicit in the

BWB method, where each up-going Bessel wave is forced to be related simply to the corresponding down-going Bessel wave.

To give the flavor of the calculation, we publish here the M1 equation associated with s-polarization:

$$\sum_{l=l_{\min}}^{l_{\max}} \frac{(-i)^l}{8\pi} \sqrt{(2l+1)(2l'+1)\frac{(l-|m|)!(l'-|m|)!}{(l+|m|)!(l'+|m|)!}}$$

$$\times \left[a_l \int_0^1 \left(1 - (-1)^{l+m} r_s(x)\right) \left(P_{l-1}^{m+1}(x) + (l+m)(l+m-1)P_{l-1}^{m-1}(x)\right) P_{l'}^m(x)\, \mathrm{d}x \right.$$

$$\left. + b_l(i) \int_0^1 \left(1 + (-1)^{l+m} r_s(x)\right) \left(P_l^{m+1}(x) - (l+m)(l-m+1)P_l^{m-1}(x)\right) P_{l'}^m(x)\, \mathrm{d}x \right] = 0.$$

$$(2.38)$$

To perform this calculation, we need to numerically perform the following integrals containing the associated Legendre functions:

$$\int_0^1 P_\mu^{m+1} P_\nu^m\, \mathrm{d}x, \qquad \int_0^1 P_\mu^{m-1} P_\nu^m\, \mathrm{d}x,$$
$$\int_0^1 P_\mu^{m+1} P_\nu^m r_s\, \mathrm{d}x, \qquad \int_0^1 P_\mu^{m-1} P_\nu^m r_s\, \mathrm{d}x,$$
$$\int_0^1 P_\mu^{m+1} P_\nu^m r_p\, \mathrm{d}x, \qquad \int_0^1 P_\mu^{m-1} P_\nu^m r_p\, \mathrm{d}x.$$

The integrals in the first row need only be evaluated once. However, because r_s and r_p are functions of k as well as $x \equiv \cos\theta_k$, the remaining (eight!) real-valued integrals must be numerically performed each time $A(k)$ is created. Consequently, Variant 2 takes a much longer time to run than Variant 1. The author has optimized the code for the computation of the integrands, allowing Variant 2 can be used as a check that Variant 1 is working well. We have found that the results of the two variants agree quite well, allowing us to use the faster method. One solution that may be both fast and elegant is to use Variant 2 but to perform the integrations on a grid of complex k at beginning of the program's execution, using an existing 2D cubic spline interpolation package. Another direction of improvement is to incorporate the modifications of the S-matrix method described in Refs. [35, 34] which expedite the search for k through approximations. These procedures have not yet been tried.

2.6 Accuracy of Results

One quickly obtained indicator of the accuracy of a result is the normalized residual, Δ_n. After the matrix A is constructed, each row of A is normalized so that the sum of the squares of its elements is 1. After the program has found a solution, we can calculate the normalized residual from the residual (Δ_r from Eq. (2.1)) of the solution.

$$\Delta_n \equiv \frac{\Delta_r}{|y| \times (\text{number of rows in } A)^{1/2}}. \tag{2.39}$$

The normalized residual has the interpretation of being the average error (RHS) of a single row equation. A random y vector would produce a normalized error of order 1. Although in principle, the solutions of $Ay = b$ should have $\Delta_n = 0$, a nonzero value is of course produced by the calculation. If Δ_n is uncharacteristically large (say 0.001), it may indicate that the truncated basis used is not sufficient to solve the resonance problem. Often, varying the seed equation, so that it "favors" a mode solution that, for the moment, has a large Δ_n, will result in a lower Δ_n and a more accurate solution. (Varying the seed equation is often a necessary step for resolving near-degenerate modes, as done in the paraxial mode mixing calculations of Chap. 3.)

It is apparent that having a low Δ_n does not by itself guarantee an accurate solution. One objection that may be raised is that this quantity only checks the wave solution at certain "points" (the locations on M2 and the finite set of equations characterizing M1). A number of tests have been done on various solutions (although few of the specific solutions presented in this thesis) that indicate that if $\Delta_n \ll 1$, the fields are very close to correct at points other than the points specified in the construction of A. To address M2, the fields at new locations along M2 were specified. To address M1, the fields in layers 0 and 1 at their interface were calculated and were seen to match the continuity equations well. An additional test verified Maxwell's equations directly in layer 0, using discrete derivatives. (Maxwell's equations should be correct even for a random vector y; this test verifies the correctness of the basis functions themselves.) Good results from these tests indicated that a small Δ_n indicated a good solution to boundary conditions of the billiard. Typical values of Δ_n for paraxial modes are 10^{-10} to 10^{-4}. Nonparaxial mode solutions often had significantly larger Δ_n. Cases in which these larger Δ_n values did not drop when the size of the finite basis was increased (and when other parameters of the numerical solution were adjusted) are most likely cases in which the (nonparaxial) modes have a significant

evanescent component.

Other difficulties can occur. If the truncated basis is too large, the solutions that are found begin to attempt to set the field to zero in the entire cavity region, with some regions of layer 0 outside the cavity having field intensities that are orders of magnitude larger than the field inside the cavity. This problem is especially prominent for the Bessel wave method, and can occur for surprisingly low values of N_{θ_k}. Yet taking N_{θ_k} to be too small often causes solutions to simply not be found: dips in the graph of Δ_r vs. $\mathrm{Re}\,k$ can simply disappear. These problems are significantly less severe for the two-basis method and therefore we most commonly use this method. Alternatively, for paraxial modes, excluding Bessel waves with θ_k larger than some intermediate angle such as $\pi/4$ can prevent some or all of the problems of the Bessel wave method.

In conclusion, the combination of both the Bessel wave method and the two-basis method works very well to find the eigenmodes and complex eigenvalues of the dome cavity. A number of demonstrations of mode calculations are given in Ref. [24]. In the following chapter we will focus on a specific set of results which demonstrate what is perhaps the most important effect of the Bragg mirror that we have discovered.

Chapter 3

SPIN-ORBIT COUPLING OF PARAXIAL MODES IN THE DOME CAVITY

This chapter discusses results that have been partially published in Refs. [26, 24]. We find that there exist true eigenmodes of the dome cavity which have cross sections that are mixtures (fixed linear combinations) of the cross sections of eigenmodes for orbital and spin angular momentum. The mixing, which occurs between pairs of modes, is characterized by a real mixing angle α. The behavior of α under modifications to the length, or other parameters of the cavity, is found to be due to the effective birefringence of the Bragg mirror. (The Bragg mirrors in our model are always made of isotropic, non-birefringent materials; the birefringence we speak of is related to the difference between r_s and r_p.) If the flat mirror is a conducting mirror ($r_s = r_p = -1$), or a virtual mirror ($r_s = r_p = +1$), the mixing angle is always 0, as these mirrors are not birefringent. For situations with large birefringence, the mixing angle approaches $\pm 45°$. The mixing angle behavior persists arbitrarily far into the paraxial limit, even though $r_s(\theta_k) - r_p(\theta_k) = 0$ at $\theta_k = 0$. Thus the mixing is zero order in small paraxial parameters, such as the spread of θ_k. At the end of this chapter we give conclusions and an outlook on observation and application of spin-orbit coupling.

It appears that our work is the first to demonstrate a mixing between spin and orbital eigenmodes. We do not as yet have a underlying Hamiltonian or Lagrangian or differential equation in which a spin-orbit interaction term is manifest: the next chapter discusses several attempts to model the cavity-with-Bragg-mirror system.

Figure 3.1: A $t = 0$ snapshot of lowest-order Gaussian beam. The envelope at $w(z)$ is shown by the thick red curves. The pink curves are parts of spheres with locations $\{z^*\}$ given by the phase condition $kz^* - \arctan(z^*/z_R) = 2\pi n$, $n \in \mathbb{Z}$, and radii $R(z^*)$. The spheres are seen to match the wavefronts well.
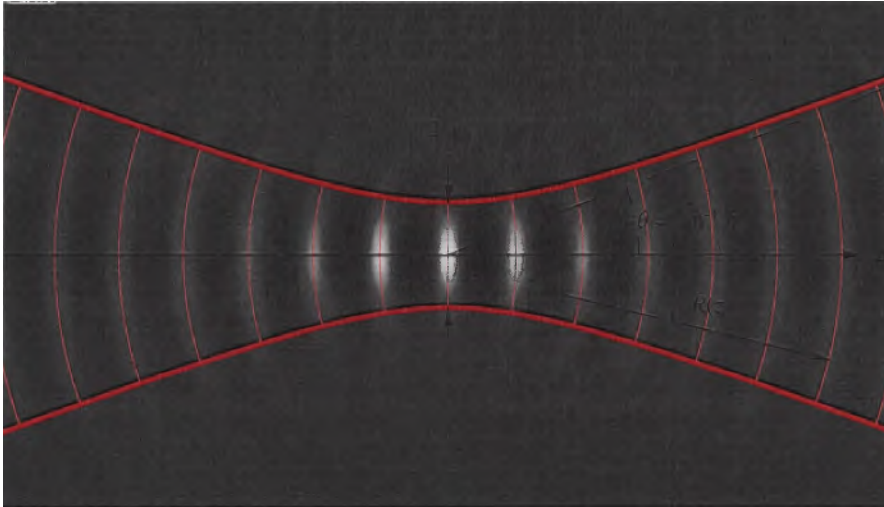
Our evidence of a spin-orbit interaction lies in both the mixing and the degeneracy splitting of the mode family (the latter of which is not new, going at least back to Erickson [36]). A handful of more direct approaches to optical spin-orbit interactions exist in the literature, none of which demonstrate non-zero mixing of paraxial modes. In their impressive development, Liberman and Zel'dovich [37], discuss a Hamiltonian formulation of polarized beams traveling through heterogeneous, isotropic media; the spin affects the trajectory of the beams, and the spin evolves along with the ray trajectory. Recently, Bliokh and Frolov S38] discuss the spin-orbit interaction and degeneracy lifting in a cylindrical dielectric resonator with periodic boundary conditions and with a possible radial variation of the refractive index. Their results do not consider or predict non-zero mixing angles.

Except where otherwise noted, the results of Sections 3.2-3.4 represent the individual, original work of the author.

3.1 Gaussian Beams, Angular Momentum, and Mode Mixing

3.1.1 The Paraxial Approximation and Gaussian Beams

The light beam from a typical laser cavity has long been known to be much like a plane wave, in that it has wavefronts that are nearly planar and nearly evenly spaced in the direction of propagation. This fact alone motivates an approximation known as the paraxial approximation, which leads to the family of Gaussian beams. The paraxial approximation for the vectorial fields is straightforwardly analyzed by an ordered expansion in [39]. In this section we will be using what amounts to the zero-order approximation (although it will appear uncontrolled because we will use the ubiquitous, hand-wavey exposition that is much shorter than the controlled version), while in Appendix B we will include the first order term, which appends nonzero longitudinal (z) vector components to the fields without modifying the transverse component fields. A discussion of the (zero-order) Gaussian beams is given in the *Lasers* books, Refs. [27, 19]. Here we will mainly be summarizing the results.

For a beam traveling along the z axis in free space, we write the fields in the form

$$\boldsymbol{E} \approx \boldsymbol{\epsilon}\psi(x,y,z,t)$$
$$= \boldsymbol{\epsilon}u(x,y,z)e^{ikz}e^{-ikct}, \tag{3.1}$$

$$\boldsymbol{H} \approx \sqrt{\epsilon/\mu_0}\,\psi(x,y,z,t)\hat{R}_z(\pi/2)\boldsymbol{\epsilon}, \tag{3.2}$$

where the envelope function u varies slowly with z, and $\boldsymbol{\epsilon}$ is the Jones vector (a unit vector in the x-y plane that may be complex). We will usually be assuming k is real in this chapter and the time dependence will be suppressed as usual. The correct equations in a medium can be obtained by replacing k with kn.

Maxwell's equations lead to the vector Helmholtz equation for $\boldsymbol{E}$,

$$\nabla^2 \boldsymbol{E} + k^2 \boldsymbol{E} = 0. \tag{3.3}$$

The approximation $\boldsymbol{E} = \boldsymbol{\epsilon}\psi$ leads to the scalar Helmholtz equation, $\nabla^2\psi + k^2\psi = 0$, which yields

$$(\partial_x^2 + \partial_y^2)u + 2ik\partial_z u + \partial_z^2 u = 0. \tag{3.4}$$

(The $k^2\psi$ term is canceled by the term in which ∂_z^2 acts on e^{ikz}.) Since u is slowly varying in z compared to e^{ikz}, $|\partial_z^2 u| \ll |k\partial_z u|$ and the third term in (3.4) can be dropped. This part of the paraxial approximation is (also) known as "the" paraxial approximation, and the resulting equation,

$$\nabla_\mathrm{T}^2 u + 2ik\partial_z u = 0, \qquad (3.5)$$

is commonly called the paraxial wave equation. This equation is a parabolic partial differential equation, the solution set of which are is the infinite family of Gaussian beams.[1] The lowest order Gaussian beam, also called the TEM$_{00}$ mode or simply "the Gaussian beam", is illustrated in Fig. 3.1. The envelope field u is given by

$$u_0 = \sqrt{\frac{\pi}{2}} \frac{1}{w(z)} \exp\left[-i\arctan\frac{z}{z_\mathrm{R}}\right] \exp\left[-\frac{x^2+y^2}{w^2(z)} + ik\frac{x^2+y^2}{2R(z)}\right], \qquad (3.6)$$

where

$$z_\mathrm{R} = \pi w_0^2/\lambda,$$

$$w(z) = w_0\sqrt{1 + \left(\frac{z}{z_\mathrm{R}}\right)^2},$$

$$R(z) = z + \frac{z_\mathrm{R}^2}{z}, \qquad (3.7)$$

with z_R being called the Rayleigh range, $w(z)$ being the $1/e$ field radius of the beam at z, and $R(z)$ being the radius of curvature of the wave fronts on the axis at z. These expressions contain two free, real parameters, which we will choose to be the wavelength λ, which sets the overall scale, and (what we will call) the paraxiality parameter, $h \equiv \lambda/(\pi w_0)$, which determines how much the beam is focused; the asymptotic divergence angle shown in Fig. 3.1 is equal to $\arctan h$. The quantity w_0, called the waist radius, gives the $1/e$ field radius of the beam at its narrowest point (the waist) which occurs at $z = 0$. (Note that w denotes a beam radius, while ω denotes angular frequency.) We note that as $h \to 0$, the beam becomes more like a plane wave and the paraxial approximation itself becomes more valid.

---

[1]Replace $k\partial_z$ with $(m/\hbar)\partial_t$ and the paraxial wave equation becomes the free-space Schrödinger equation in 2D. The solutions given here for the envelope function u are also the solutions for the time evolving wavefunction in 2D space. Another, more involved transformation, results in a 1D harmonic oscillator Schrödinger equation. For a discussion on the equivalence of paraxial beams to the quantum harmonic oscillator, see Ref. [40].

3.1.2   Higher Order Gaussian Modes and Gaussian Modes in Cavities

Eq. (3.6) is not the only solution to the paraxial wave equation. For any fixed values of λ and h, there exists a countably infinite set of independent solutions. We will refer to these solutions, as well as the one in (3.6) as modes instead of beams. This set of vector modes can be written in two bases, the Hermite-Gauss basis and the Laguerre-Gauss basis. The Hermite-Gauss modes are given by

$$
\mathrm{HG}_{mn}(x,y,z) = \sqrt{\frac{2}{2^{m+n}m!n!\pi}}\frac{1}{w(z)}\exp\Big(ikz - i(m+n+1)\arctan\frac{z}{z_{\mathrm{R}}}\Big)
$$
$$
\times H_m\Big(\frac{\sqrt{2}x}{w(z)}\Big)H_n\Big(\frac{\sqrt{2}y}{w(z)}\Big)\exp\Big(-\frac{x^2+y^2}{w^2(z)} + ik\frac{x^2+y^2}{2R(z)}\Big), \qquad (3.8)
$$

where H_m denotes the Hermite polynomial, and $n,m \in \{0,1,\dots\}$. The Laguerre-Gauss modes are given by

$$
\mathrm{LG}_p^l(\rho,\phi,z) = \sqrt{\frac{2}{\pi}\frac{p!}{(p+|l|)!}}\frac{1}{w(z)}\exp\Big(ikz - i(2p+|l|+1)\arctan\frac{z}{z_{\mathrm{R}}}\Big)
$$
$$
\times \Big(\frac{\sqrt{2}\rho}{w(z)}\Big)^{|l|}L_p^{|l|}\Big(\frac{2\rho^2}{w^2(z)}\Big)\exp\Big(-\frac{\rho^2}{w^2(z)} + ik\frac{\rho^2}{2R(z)}\Big)\exp\Big(il\phi\Big), \qquad (3.9)
$$
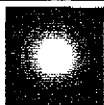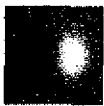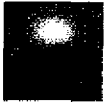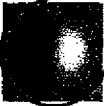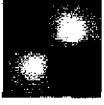
where L_p^l denotes the generalized Laguerre functions, and $p \in \{0,1,\dots\}$ and $l \in \{-p,-p+1,\dots,p\}$. The most important aspect of the Laguerre-Gauss modes is the $e^{il\phi}$ factor. For any fixed z, the sets $\{\mathrm{HG}_{mn}\}$ and $\{\mathrm{LG}_p^l\}$ both form complete, orthogonal bases for scalar functions of x and y [27]. The HG and LG functions are normalized according to

$$
\iint \mathrm{HG}_{nm}^* \mathrm{HG}_{n'm'}\,\mathrm{d}x\mathrm{d}y = \delta_{nn'}\delta_{mm'},
$$
$$
\iint (\mathrm{LG}_p^l)^* \mathrm{LG}_{p'}^{l'}\,\mathrm{d}x\mathrm{d}y = \delta_{pp'}\delta_{ll'}, \qquad (3.10)
$$

for any fixed z.

The HG and LG families are each subdivided into groups of the same transverse order, N. (This N should not be confused with the N representing the number of layers in a Bragg mirror.) For the HG family, $N = n + m$ and for the LG family, $N = 2p + |l|$. For both families, there is a single $N = 0$ mode corresponding to

Table 3.1: Cross sections of the Gaussian modes (scalar fields).

	HG: n, m		LG: p, l	
$N = 0$		$0, 0$		$0, 0$
$N = 1$		$1, 0$		$0, 1$
		$0, 1$		$0, -1$
$N = 2$		$1, 1$		$1, 0$
		$2, 0$		$0, 2$
		$0, 2$		$0, -2$

the TEM$_{00}$ mode, and for both families, the number of modes with order N is seen to be $N + 1$. Table 3.1 shows the cross section of the HG and LG modes up to $N = 2$. The curved arrows indicate that the entire wave pattern *rotates at optical frequencies*. These rotating wave patterns can be turned into stationary patterns by superimposing two modes with opposite signs of l. The set of HG modes of transverse order N is related to the set of LG modes of order N by a known unitary transformation [41, 42]. Typically, the expansion of an HG mode pattern in LG mode patterns involves all of the LG mode patterns in the given family (for example $\mathrm{HG}_{20} = (1/2)(\mathrm{LG}_0^2 + \mathrm{LG}_0^{-2} - \sqrt{2}\mathrm{LG}_1^0)$). In contrast, the allowed mixtures of LG mode patterns which we discuss in this chapter only involve two scalar LG mode patterns (and, in one sense, four vector LG mode patterns). The superscripts and subscripts in the HG_{nm} and LG_p^l notation correspond to numbers of nodes in the transverse wave patterns. The number n and m correspond to the numbers of vertical and horizontal nodes, respectively. The number p corresponds to the number of radial nodes at finite,

nonzero radius. The number $|l|$ corresponds to the number of diametrical nodal lines.

For a mode of transverse order N, the shape of the wavefronts is seen from both (3.8) and (3.9) to be given by

$$\varphi = kz - (N+1)\arctan\left(\frac{z}{z_R}\right) + \frac{k\rho^2}{2R(z)} = \text{const.} \qquad (3.11)$$

This shape, near the axis, is approximated by a parabola or a sphere.[2] Fig. 3.1 demonstrates that appropriately placed spheres are good approximations to the Gaussian mode wavefronts over the angular region where the field has significant magnitude. If a counterpropagating beam is added to the original Gaussian, the φ dependence of the field changes from the traveling wave $e^{i(\varphi-\omega t)}$ to the standing wave $\cos(\varphi)e^{-i\omega t}$. Two spherical conducting mirrors may be placed anywhere that $\cos\varphi = 0$ (where the field is zero), and we see that the Gaussian standing wave is the (approximate) eigenfunction of the cavity formed by the two mirrors. Thus standing Gaussian modes are the modes of cavities of the shape we are interested in. For reference we can generalize (3.11) to the standing wave condition:

$$k(z_2 - z_1) - (N+1)\left(\arctan\frac{z_2 - z_\text{w}}{z_\text{R}} + \arctan\frac{z_\text{w} - z_1}{z_\text{R}}\right) = (q+1)\pi - \frac{\phi_1}{2} - \frac{\phi_2}{2}. \qquad (3.12)$$

Here z_1 and $z_2 > z_1$ are the positions of the mirrors, which have reflection phase shifts (for normal incidence) of ϕ_1 and ϕ_2. The beam waist is located at z_w and q is an integer. In the context of cavity modes, the partition of the Gaussian families by transverse order N suddenly makes sense: the $N+1$ modes of order N are, in paraxial approximation, *degenerate*, having the same k (because only N enters Eqs. (3.11, 3.12), not n or m or l or p).

The physical situation is as follows. Two conducting mirrors are placed to form a cavity for which the axial bouncing-ball trajectory is ray-stable. The cavity shape fixes (quantizes) all of the Gaussian mode parameters, including the location of the beam waist. Thus in the paraxial approximation, the eigenmodes come in degenerate sets, each set denoted by the integer pair (q, N), where q is the longitudinal order,

---

[2]Taking the limit $z \to 0$ leads to a parabolic wavefront. Contrary to a common misconception, the $z \to \infty$ limit does not lead to spherical wavefronts. (Although spheres are a better approximation than paraboloids, the ρ^4 coefficient is still incorrect.) The $z \to \infty$ limit actually yields oblate ellipsoidal wavefronts, with the ellipsoid centered halfway between the origin and the wavefront, and the x and y semimajor axes being $\sqrt{2}$ times larger than the z semimajor axis.

roughly the number of longitudinal antinodes in the modes. In this approximation, the eigenmodes can be any linear transformation of the Gaussian modes with the same q and N.

3.1.3 Classification and Mixing of the Vector Laguerre-Gauss Modes

Including the two types of polarization causes the mode degeneracy to become of order $2(N+1)$. Using the LG basis for the scalar fields and the circular polarization basis $\{\sigma^+, \sigma^-\}$, the expansion of the electric field becomes

$$
\begin{pmatrix} E_x \\ E_y \end{pmatrix} \approx \sum_{\substack{p=0,1,\ldots,\lfloor N/2 \rfloor \\ l=-N,-N+2,\ldots,N}}^{2p+|l|=N} \left[A_p^l \underbrace{\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}}_{\sigma^+} + B_p^l \underbrace{\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}}_{\sigma^-} \right] \mathrm{LG}_p^l(\rho, \phi), \qquad (3.13)
$$

where the constraint above the sum must always be satisfied, and A_p^l and B_p^l are the complex expansion coefficients. The z-dependence changes (significantly) only in overall phase as the nature or separation of the mirrors changes. Henceforth, the notation LG_p^l will refer either to a true mode (complete with z dependence and a definite frequency), or to the cross section *mode pattern* (a function of ρ and ϕ only) obtained from setting $z = 0$ in Eq. (3.9):

$$
\mathrm{LG}_p^l(\rho, \phi) = \sqrt{\frac{2p!}{\pi(p+|l|)!}} \frac{1}{w_0} \left(\frac{\sqrt{2}\rho}{w_0} \right)^{|l|} L_p^{|l|} \left(\frac{2\rho^2}{w_0^2} \right) \exp\left[-\frac{\rho^2}{w_0^2} + il\phi \right]. \qquad (3.14)
$$

The field expansion (3.13) is given in the vector Laguerre-Gauss basis, with the basis functions denoted by $\mathrm{LG}_p^l \sigma^s$. Here s can be interpreted as the spin and is allowed to be ± 1, denoted simply by "+" or "−" when typeset as a superscript. The value l can be interpreted as the orbital angular momentum, and the total angular momentum is $m = l + s$. The vector LG basis states are simultaneous eigenstates of total, orbital, and spin angular momentum. The interpretation of m, l, and s as angular momenta is discussed in Appendix B and the references cited therein. Here we will proceed without needing the physical interpretation, showing that $m = l + s$ by a comparison of terms.

Any paraxial mode in the narrow frequency range corresponding to (q, N) should be of the form given in Eq. (3.13). Since the degeneracy of the modes with the

same transverse order is derived in the paraxial approximation, we expect that this degeneracy is, in reality, slightly broken.[3] Thus the summation will be partitioned into several summations so that the total number of degrees of freedom, $2(N+1)$, remains unchanged. Now we assume that we can select the quantum number m. In Sec. 2.3 it was proven, without any need to link m with the physical quantity of angular momentum, that the true eigenmodes of the system can be chosen to also be modes of definite m. Thus we can choose one value of m at a time (later in this section we establish the rules for mixing and superposition, which govern which m values can be present in a single mode). Thus, independent of anything to do with the paraxial mode expansion (3.13), it must be the case that $\boldsymbol{E}_{\mathrm{T}} \equiv \boldsymbol{E} - E_z \hat{z}$ for the eigenmodes can be written as $\boldsymbol{E}_{\mathrm{T}} = E_\rho \hat{\rho} + E_\phi \hat{\phi}$, where E_ρ and E_ϕ have a sole ϕ dependence of $e^{im\phi}$. From the Cartesian components of $\hat{\rho}$, and $\hat{\phi}$, we get

$$\boldsymbol{E}_{\mathrm{T}} = \frac{1}{\sqrt{2}} \underbrace{(E_\rho + iE_\phi)}_{\propto \exp(im\phi)} e^{i\phi} \sigma^- + \frac{1}{\sqrt{2}} \underbrace{(E_\rho - iE_\phi)}_{\propto \exp(im\phi)} e^{-i\phi} \sigma^+. \tag{3.15}$$

Comparing (3.15) with (3.13) reveals that at most two terms in (3.13) are present: the $A^{m-1}_{(N-|m-1|)/2}$ term can be present if $(N - |m-1|)/2$ is in $\{0, 1, \ldots, \lfloor N/2 \rfloor\}$, the range of p, and the $B^{m+1}_{(N-|m+1|)/2}$ term can also be present if $(N - |m+1|)/2$ is in the range of p. If both terms are present and $m \neq 0$, these two pure Laguerre-Gauss mode patterns mix to form the cross sections of two nearly degenerate eigenstates. (If both terms are present and $m = 0$, the two pure LG states are *exactly* degenerate[4], and an arbitrary (excitation-dependent) superposition can be made.) Thus, we have isolated pairs of modes that are mixable, and we will find that the nature of the stack determines the nature of the mixing. Note that there is no general argument from paraxial theory that says the pure LG mode patterns are preferred over mixed patterns. Thus we see that the degree to which modes are mixed is zero order in h; it does not go to zero as h goes to zero. It is important to realize that, if the

---

[3]Ref. [43] claims that the degeneracy is not broken when M1 is a planar conductor and M2 is a spherical conductor, in contradiction with our numerical findings. This reference however, assumes from the beginning that $E_y = 0$, an assumption that is not justified. The breaking of the transverse degeneracy in conducting cavities was both predicted and observed in Refs. [44, 45, 46, 47], which we will discuss later, nearly two decades before Ref. [43].

[4]At first thought, it is hardly believable that the modes $\mathrm{LG}^l_p \sigma^+$ and $\mathrm{LG}^l_p \sigma^-$ are not degenerate in general. This becomes believable when one plots the instantaneous vector fields and finds that they are completely different for these two modes. The mode $\mathrm{LG}^2_0 \sigma^-$ (mode B) is plotted in Fig. 3.3, while the mode $\mathrm{LG}^2_0 \sigma^+$ is plotted in Fig. 7 of Ref. [25]. An analytic expression for the frequency difference, in leading order in h, of these two modes is derived in Ref. [47, 46].

Table 3.2: The vector LG modes.

N	m	l	E_T	N	m	l	E_T
0	-1	0	$LG_0^0 \sigma^-$	3	-4	-3	$LG_0^{-3} \sigma^-$
	1	0	$LG_0^0 \sigma^+$		-2	-3	$LG_0^{-3} \sigma^+$
					-2	-1	$LG_1^{-1} \sigma^-$
1	-2	-1	$LG_0^{-1} \sigma^-$		0	-1	$LG_1^{-1} \sigma^+$
	0	-1	$LG_0^{-1} \sigma^+$		0	1	$LG_1^1 \sigma^-$
	0	1	$LG_0^1 \sigma^-$		2	1	$LG_1^1 \sigma^+$
	2	1	$LG_0^1 \sigma^+$		2	3	$LG_0^3 \sigma^-$
					4	3	$LG_0^3 \sigma^+$
2	-3	-2	$LG_0^{-2} \sigma^-$				
	-1	-2	$LG_0^{-2} \sigma^+$				
	-1	0	$LG_1^0 \sigma^-$				
	1	0	A $LG_1^0 \sigma^+$				
	1	2	B $LG_0^2 \sigma^-$				
	3	2	$LG_0^2 \sigma^+$				

linear cavity problem is formulated with a *scalar field* instead of the correct vector electromagnetic field, the equation analogous to (3.15) has only one term, not two. Thus there is no possibility of mixing (no pairs) in the scalar problem, regardless of whether the action of M1 is simple (r = const.) or complicated ($r = r(\theta_k)$). This fact corroborates with our numerical results (Sec. 3.3) that $r_s - r_p$ is the important quantity for mixing; this quantity does not exist in the scalar problem, which has only one reflection function r (which turns out to be the same as function r_s).

Table 3.2 shows the vector LG modes/mode patterns through $N = 3$. The brackets in the rightmost columns will be explained later in this section. The dashed boxes denote the mixable pairs, with two mixable mode patterns being labeled A and B for future reference. Mixable pairs occur for all orders $N \geq 2$. For fixed $m = +1$, we can say that the nature of the cavity, including the dielectric stack, mixes A and B. Labeling the (cross sections of) the new, normalized modes as C and D, the mixing

should be described by the two-parameter special unitary transformation[5]

$$\begin{pmatrix} C \\ D \end{pmatrix} \approx \begin{pmatrix} \cos\alpha & -e^{i\beta}\sin\alpha \\ e^{-i\beta}\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix}. \tag{3.16}$$

$(C \cdot D = 0$ yields 2 real equations (constraints), the fact that we want C and D to be normalized yields 2 constraints, and the fact that we don't care about the phases of C and D yields 2 constraints. There are initially 8 parameters in the matrix; subtracting the constraints yields two free parameters, α and β.) This prediction includes the hypothesis that, analogous to quantum mechanical degenerate perturbation theory, the new modes will indeed be orthogonal. As will be presented in Sec. 3.2, the numerical results show that even the transformation (3.16) is too general; it appears from the calculations that $\beta = 0$, causing the transformation to simply be the orthogonal 2D rotation $\hat{R}(\alpha)$. The reason for the loss of β has not yet been determined. *We will use the convention that the mode with the larger* Re k *is always labeled D.* With this convention, the range of α can be taken to be $[-45°, 45°]$.

The brackets on the right side of Table 3.2 link pairs of modes that are *exactly* degenerate because they are related by symmetry of the cavity. The vector fields $\mathrm{LG}_p^l\sigma^+$ and $\mathrm{LG}_p^{-l}\sigma^-$ are related by reflection about the y axis (the x axis also works). That is, taking $l \to -l$ and $s \to -s$ amounts to a reflection. This symmetry indicates that there should be an exact degeneracy that allows arbitrary superpositions to be made. The nature of this set of arbitrary positions can be seen by the following argument. Temporarily label the mode pattern $\mathrm{LG}_1^0\sigma^-$ as W and the mode pattern $\mathrm{LG}_0^{-2}\sigma^+$ as X. Imagine the cavity being excited by a $m = -1$ beam. By symmetry we see that the new mixed modes Y and Z must be related to W and X by the same transformation that relates C and D to A and B. That is,

$$\begin{pmatrix} C \\ D \end{pmatrix} \approx \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix},$$

$$\begin{pmatrix} Y \\ Z \end{pmatrix} \approx \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} W \\ X \end{pmatrix}. \tag{3.17}$$

---

[5]Our matrix has one less free parameter than a general $D^{(1/2)}$ matrix (or a general $SU(2)$ matrix), because the D^j matrices must account for full rigid body rotations by the three Euler angles. Replacing α with $\theta/2$ and β with ϕ in our matrix gives the $D^{(1/2)}$ matrix which rotates a spinor initially pointing positively along the z direction in real space to a spinor pointing positively along the direction $(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ in real space.

Not only do we expect the mixtures to be the same, but we expect the exact equalities $k_Y = k_C$ and $k_Z = k_D$. Thus we may arbitrarily superpose the mixed modes Y and C (or Z and D) by exciting with an input beam that is a superposition of $m = 1$ and $m = -1$ beams. The superposition of these mixed modes is equivalent to a mixing of superposed modes. This can be seen by considering the superposition $V = aY + bC$ with a and b being in general complex. V is given by

$$V = a\cos(\alpha)W - a\sin(\alpha)X + b\cos(\alpha)A - b\sin(\alpha)B$$
$$= \cos(\alpha)(aW + bA) - \sin(\alpha)(aX + bB). \tag{3.18}$$

In this interpretation, the excitation-selected, unperturbed modes $(aW + bA)$ and $(aX + bB)$, are mixed by the cavity. At this point the reader may wonder, "What if the input beam is a superposition of $m = 1$ and, say, $m = 3$?" The answer is that, in general, a *superposition* of m and m' where $|m'| \neq |m|$ will not be allowed (in the limit of negligible resonance widths) because generally there is no exact degeneracy involved. Furthermore, a *mixture* of m and m' where $m' \neq m$ is not allowed by the argument of Sec. 2.3. We note that mixing between modes of the same m and different N (and generally different l and s as well) can occur, but because these modes are not degenerate in paraxial theory, the mixing amplitude will be first order, not zero order, in h, and thus it will be usually negligible compared to the mixing amplitude α which we are considering. To summarize: **the $2(N+1)$-fold degeneracy (of the vector cavity modes specified by (q, N)) is in general partitioned into $N+1$ different 2-fold exact degeneracies with eigenfrequencies** $\omega_1, \omega_2, \ldots, \omega_{N+1}$. The superposition chosen within each exactly degenerate pair is dependent on how the cavity is excited. There are $\lfloor N/2 \rfloor$ mixable pairs *of* exactly degenerate pairs and hence there are $\lfloor N/2 \rfloor$ different mixing angles α.

The two most important points of this section are: (1) the cylindrically symmetric cavity is free to mix certain pairs of modes having different orbital and spin angular momentum, but the same total angular momentum (we can refer to this as classical spin-orbit coupling); (2) the amplitude of the mixing, α, is zero order in h, allowing large mixing angles to persist in the paraxial limit. We note that the difference in wave number of the mixed modes, for example $k_D - k_C$, is first order (or some positive order) in h, and is therefore small. If the resonance widths of C and D are as large as or larger than $k_D - k_C$, the mixing cannot be observed; it will appear that arbitrary superpositions of A and B can be made. It is this requirement which accounts for the

fact that this splitting has not yet been experimentally observed. Before proceeding to the mixing angle results, we will derive the Bessel wave decomposition for LG beams and for mixed modes. In Appendix B we relate the quantum numbers m, l, and s to the physical angular momentum quantities of optical beams and modes.

3.1.4 Bessel Wave Decomposition of LG Beams and Modes

We wish to determine the Bessel wave decomposition of the electromagnetic field when

$$E_{\mathrm{T}}(\rho, \phi, 0) = \mathrm{LG}_p^l(k, w_0; \rho, \phi)\boldsymbol{\sigma}^s, \qquad (3.19)$$

for fixed m, N and s (l and p are dependent on these). To simpify the expressions we take $z = 0$ to be the location of the free-space waist location of the mode envelopes. (If the actual waist lies inside the dielectric mirror, the free-space waist location is the z position that the waist would be at if the envelope of the mode in layer 0 were continued into the stack region as if it were free space.) For the numerical cavity problem, this means the coordinate system we are now using has been shifted in z, since the numerical cavity problem cannot generally be formulated ahead of time such that the origin is located where the free-space waist of the mode will be. The (new) value of z_1 now gives the location of the M1 surface relative to the mode waist, while the new value z_d gives the location of the zenith of the M2 dome relative to the mode waist.) Note that if there is a longitudinal node at $z = 0$, so that $E_{\mathrm{T}}(\rho, \phi, 0) = 0$, the decomposition procedure must be modified, perhaps by setting $H_{\mathrm{T}}(\rho, \phi, 0) = \mathrm{LG}_p^l(k, w_0; \rho, \phi)\boldsymbol{\sigma}^s$. We assume a refractive index of 1 for layer 0 (to change this in this section, replace k with kn and deal with h in a consistent way). We develop the solution which will be valid for traveling beams (no mirrors), for standing waves formed by an LG beam reflected off a dielectric planar mirror (no M2 involved), and for cavity modes which have parameters set so that $\alpha = 0$. The direction variable θ_k is now restricted to the range $[0, \pi/2)$ and directional coefficients s_{u}, s_{d}, p_{u}, and p_{d}, are used. The "u" and "d" subscripts stand for up and down (up is in the positive z direction). For a LG *beam* moving in the positive z direction in free space, we simply have $s_{\mathrm{u}} = p_{\mathrm{u}} = 1$ and $s_{\mathrm{d}} = p_{\mathrm{d}} = 0$. However, for a standing wave formed by a focused beam being reflected off of dielectric mirror (whether there

exists an M2 mirror or not) the coefficients can be set[6] to be

$$s_{\mathrm{d}} = p_{\mathrm{d}} = e^{ikz_1\cos\theta_k},$$

$$s_{\mathrm{u}} = e^{-ikz_1\cos\theta_k} r_{\mathrm{s}}(k,\theta_k),$$

$$p_{\mathrm{u}} = e^{-ikz_1\cos\theta_k} r_{\mathrm{p}}(k,\theta_k). \tag{3.20}$$

Equations (2.19, 2.20, 2.21) at $z = 0$ yield,

$$E_\pm = \sqrt{2}\pi i^m e^{i(m\mp 1)\phi} \int_0^{\pi/2} \mathrm{d}\theta_k \sin(\theta_k) J_{m\mp 1}(k\rho\sin\theta_k)$$

$$\times \left[(s_{\mathrm{u}} + s_{\mathrm{d}})\bar{S}_k(\theta_k) \mp \cos(\theta_k)(p_{\mathrm{u}} + p_{\mathrm{d}})\bar{P}_k(\theta_k)\right], \tag{3.21}$$

where $\bar{S}_k(\theta_k)$ and $\bar{P}_k(\theta_k)$ are functions on $[0, \pi/2)$ so that $s_{\mathrm{u}}\bar{S}(\theta_k) = S_k(\theta_k)$, $s_{\mathrm{d}}\bar{S}(\theta_k) = S_k(\pi - \theta_k)$ and the same relations hold for p polarization.

On the other hand, we wish to enforce the correct LG mode pattern. Picking $s = +1$ sets $l = l_+ \equiv m - 1$, $p = p_+ \equiv (N - |l_+|)/2$ and

$$E_- = 0,$$

$$E_+ = \mathrm{LG}_{p_+}^{l_+}(\rho, \phi)$$

$$= \sqrt{\frac{2p_+!}{\pi(p_+ + |l_+|)!}}\frac{1}{w_0}\left(\frac{\sqrt{2}\rho}{w_0}\right)^{|l_+|} L_{p_+}^{|l_+|}\left(\frac{2\rho^2}{w_0^2}\right) e^{-\rho^2/w_0^2} \tag{3.22}$$

Likewise, picking $s = -1$ sets $l = l_- \equiv m + 1$, $p = p_- \equiv (N - |m + 1|)/2$, and

$$E_+ = 0,$$

$$E_- = \mathrm{LG}_{p_-}^{l_-}(\rho, \phi). \tag{3.23}$$

Changing variables to $x \equiv kw_0\sin(\theta_k)/\sqrt{2} = \sqrt{2}\sin(\theta_k)/h$ changes the upper limit of integration in (3.21) to $\sqrt{2}/h$. (Here we are taking k to be real.) We shall find solutions to $\bar{S}_k$ and $\bar{P}_k$ that are proportional to $e^{-x^2/2}$ so that, in the paraxial limit,

---

[6]There is some degree of arbitrariness here. Another natural definition is to set $s_{\mathrm{u}} = p_{\mathrm{u}} = 1$, $s_{\mathrm{d}} = \exp[i2kz_1\cos\theta_k]/r_{\mathrm{s}}(k,\theta_k)$, and $p_{\mathrm{d}} = \exp[i2kz_1\cos\theta_k]/r_{\mathrm{p}}(k,\theta_k)$. The ratios of up to down coefficients must be $\exp[-i2kz_1\cos\theta_k]r_{\mathrm{s/p}}$.

it is self-consistent to replace $\sqrt{2}/h$ with ∞. The result of the transformation[7] is

$$E_\pm = i^m \frac{2\sqrt{2}\pi}{k^2 w_0^2} e^{i(m\mp 1)\phi} \int_0^\infty \mathrm{d}x\, x J_{m\mp 1}[(\sqrt{2}/w_0)\rho x]$$
$$\times \left[\frac{1}{\sqrt{1 - \frac{2}{k^2 w_0^2} x^2}} (s_\mathrm{u} + s_\mathrm{d}) \bar{S}_k \mp i(p_\mathrm{u} + p_\mathrm{d}) \bar{P}_k \right]. \tag{3.24}$$

For $s = \pm 1$ we require that $E_\mp = 0$; one way to ensure this is to set

$$\bar{S}_k = \mp i \sqrt{1 - \frac{2}{k^2 w_0^2} x^2} \frac{p_\mathrm{u} + p_\mathrm{d}}{s_\mathrm{u} + s_\mathrm{d}} \bar{P}_k. \tag{3.25}$$

Now we wish to use the relation

$$\int_0^\infty J_\nu(yx) x^{\nu+1} L_n^\nu(x^2) e^{-x^2/2}\mathrm{d}x = (-1)^n y^\nu L_n^\nu(y^2) e^{-y^2/2}, \tag{3.26}$$

which comes from Eq. 7.421.4 of [48]. Comparing integrands with (3.24) and using (3.25), yields, for $s = \pm 1$:

$$\bar{P}_k = \pm(-i)^{(m-1+2p_\pm)} \frac{\zeta_\pm}{4\pi^{3/2}} \frac{k^2 w_0}{p_\mathrm{u} + p_\mathrm{d}} \sqrt{\frac{p_\pm!}{(p_\pm + |l_\pm|)!}} x^{|l_\pm|} L_{p_\pm}^{|l_\pm|}(x^2) e^{-x^2/2}, \tag{3.27}$$

where, due to the relation $J_{-n} = (-1)^n J_n$,

$$\zeta_\pm = \begin{cases} 1, & l_\pm \geq 0 \\ (-1)^{l_\pm}, & l_\pm < 0 \end{cases}. \tag{3.28}$$

In Eqs. (3.25 - 3.28), the top sign is chosen if we wish to make the $\mathrm{LG}_{p+}^{l+} \sigma^+$ mode (or beam) and the bottom sign is chosen if we wish to make the $\mathrm{LG}_{p-}^{l-} \sigma^-$ mode. Only a single approximation has been made: the approximation of the upper limit $\sqrt{2}/h$ as infinity. To the extent that this approximation is valid, we have found in (3.27) and (3.25) a set of vector Bessel waves that produces the desired field E_T at $z = 0$. The E_z field can be determined by Eq. (2.19). Furthermore, evaluation at $z \neq 0$ can be accomplished by including the phases $e^{ikz\cos\theta_k(x)}$. The full expansions of $E(\rho, \phi, z)$

---

[7]To include the z-dependence in Eq. (3.24), simply include the forward and backward propagation phases by replacing the "$(s_\mathrm{u} + s_\mathrm{d})\bar{S}_k$" with $(s_\mathrm{u} e^{ikc(x)z} + s_\mathrm{d} e^{-ikc(x)z})\bar{S}_k$ and the "$(p_\mathrm{u} + p_\mathrm{d})\bar{P}_k$" with $(p_\mathrm{u} e^{ikc(x)z} + p_\mathrm{d} e^{-ikc(x)z})\bar{P}_k$, where $c(x) = (1 - 2x^2/(k^2 w_0^2))^{(1/2)}$. See Appendix A for further z-dependent expressions.

and $\boldsymbol{H}(\rho, \phi, z)$ are given in Appendix A.

This Bessel wave decomposition may not appear to be unique, because integrands were set equal to each other, both in enforcing $E_{\mp} = 0$ and in enforcing $E_{\pm} = \mathrm{LG}_{p\pm}^{l\pm}$. However, the equations that we are enforcing must be true for all ρ, and ρ enters in the argument of the Bessel function in Eq. (3.24). We believe this ensures the uniqueness of the derived decomposition.

With the Bessel wave decompositions of the pure LGσ modes, one can attempt to combine them to give the decompositions of mixed modes. This endeavor is the subject of Sec. 4.1. The exactly degenerate pairs may also be combined (without caveat or difficulty) to give the unmixed modes with multidirectional linear polarization or, more generally, to give the unmixed modes with multidirectional elliptical polarization. In general, both of these pairwise combinations (e.i. both mixing and superposition) occur at the same time, involving four pure circularly polarized LG modes.

The form of the Bessel wave decomposition is essentially the same as the the real space LG form given in Eq. (3.14). For visual reference, the squared, circumference-weighted p polarized plane wave distribution for the LG_1^0 and LG_0^2 beams/modes are shown in Fig. 3.2. The normalized, plotted distribution is

$$\tilde{g}(\theta_k) \equiv \frac{g(\theta_k)}{\int g(\theta_k')\, \mathrm{d}\theta_k'}, \tag{3.29}$$

where

$$g(\theta_k) = \sin(\theta_k)\frac{p!}{(p+l)!}\left[(\sqrt{2}\sin(\theta_k)/h)^l L_p^l(2\sin^2(\theta_k)/h^2)\exp(-\sin^2\theta_k/h^2)\right]^2, \tag{3.30}$$

with $h = 0.065$. With the normalized distribution we can obtain the mean polar angle,

$$\bar{\theta}_k = \int \theta_k \tilde{g}(\theta_k)\, \mathrm{d}\theta_k. \tag{3.31}$$

In the limit $h \to 0$, $\sin\theta_k$ can be replaced by θ_k and one obtains

$$\lim_{h\to 0}\frac{\bar{\theta}_k}{h} = \begin{cases} \frac{7}{8}\sqrt{\pi/2} \approx 1.097 & p=1, l=0 \\ \frac{15}{16}\sqrt{\pi/2} \approx 1.175 & p=0, l=2 \end{cases}. \tag{3.32}$$

Figure 3.2: Plots of the polar angle distribution $\bar{g}(\theta_k)$ for p polarization, using $h = 0.065$, (x axis is θ_k in radians). Solid lines correspond to $p = 1$ and $l = 0$ (mode pattern A), dashed lines to $p = 0$ and $l = 2$ (mode pattern B). The s polarized distributions are similar, varying only by the factor $\cos^2 \theta_k$. The vertical lines show the mean polar angles $\bar{\theta}_k = 0.07147$ and $\bar{\theta}_k = 0.07651$ for the two distributions.
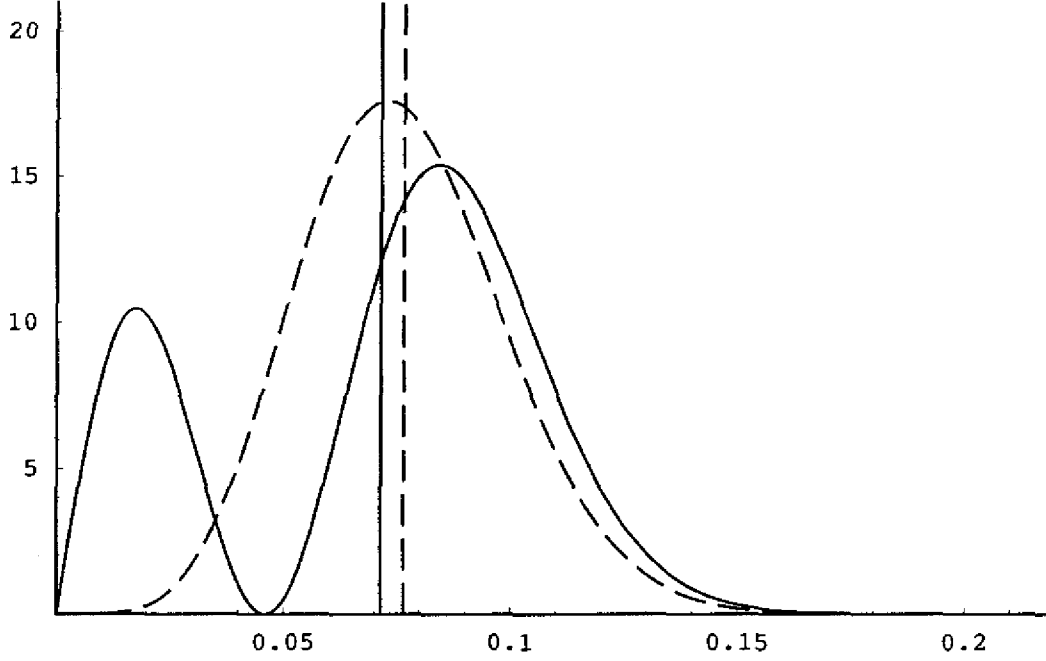
In principle, the average polar angle, $\bar{\theta}_k$, can be used as an effective angle in model theories describing mixing.

Finally in this section we discuss the relation of any paraxial Bessel wave basis expansion of $\boldsymbol{E}(x, y, z)$, including of course, the analytic BWB expansion for a single m and N derived above, to the paraxial approximation. Any expansion in Bessel or plane waves satisfies the Maxwell equations exactly, and explicitly looking at how paraxial this field is closes the following procedural loop: paraxial theory $\rightarrow$ LG pattern in $z = 0$ plane $\rightarrow$ S_k and P_k $\rightarrow$ Maxwell-satisfying Bessel wave expansion of $\boldsymbol{E}$ in all space $\rightarrow$ paraxial theory. Recall from Sec. 3.1.1 that, to zero order in h, $\boldsymbol{E} = \epsilon\psi$, with $u \equiv \psi/e^{ikz}$ satisfying the paraxial wave equation. Taking the Jones vector ϵ to be σ^+ or σ^- means that $E_\pm$ must (approximately) satisfy the paraxial wave equation. We can see that E_+ and E_- for a forward traveling beam are integral

expansions in the scalar Bessel wave form:

$$f_{\mathrm{BW}}(\boldsymbol{x}) = \underbrace{e^{in\phi} J_n(k\rho \sin\theta_k) e^{ik[\cos(\theta_k)-1]}}_{u_{\mathrm{BW}}(\boldsymbol{x})} \times e^{ikz}. \tag{3.33}$$

Since the integrand is known to decay as $e^{-(1/h^2)\sin^2\theta_k}$, it is sensible to approximate the $\sin\theta_k$ appearing in f_{BW} in Eq. (3.33) as θ_k, and to approximate the $\cos\theta_k$ appearing in f_{BW} as $(1-\theta_k^2/2)$. The resulting approximate (paraxial) envelope function for f_{BW},

$$u_{\mathrm{par.\ BW}} \equiv e^{in\phi} J_n(k\theta_k\rho) e^{ik(-\theta_k^2/2)z}, \tag{3.34}$$

is, by direct substitution, found to *exactly* solve the paraxial wave equation

$$\left[\nabla_{\mathrm{T}}^2 + 2ik\frac{\partial}{\partial z}\right] u_{\mathrm{par.\ BW}} = 0. \tag{3.35}$$

The direct substitution is straightforward with the use of the Bessel differential equation, and will not be shown here. (An even simpler way to show this is to note that the Bessel waves are integral expansions in plane waves, and a single plane wave with $u \propto e^{ik[-(1/2)\theta_k^2 z + \theta_k x]}$ satisfies the paraxial wave equation.)

This nice result works for backward propagating beams as well (there the paraxial wave equation is $[\nabla_{\mathrm{T}}^2 - 2ik\partial_z]u = 0$), and hence it works for the cavity mode solutions. The expansion of $\boldsymbol{E}$ in the (modified, $O(\theta_k^2)$) Bessel waves not only satisfies the zero-order expansion (in $h/2$) given in Lax *et al.* [39], it also satisfies the first-order expansion. Lax *et al.* show that the first order contribution is entirely given by an E_z component, which is proportional to h and satisfies its own paraxial wave equation: $[\nabla_{\mathrm{T}}^2 + 2ik\partial_z](E_z/e^{ikz}) = 0$. To show that the first order expansion is satisfied by the approximate Bessel waves, note that the expression for E_z in (2.19) is also an integral expansion in f_{BW}, and thus, by the arguments above, its envelope satisfies the paraxial wave equation. The fact that $E_z \propto h$ comes from the extra $(\sin\theta_k)$ factor present in the expression for E_z but not in the expressions for $E_\pm$. (Upon conversion to the integration variable $x = (\sqrt{2}/h)\sin\theta_k$, an extra factor of h appears in front.) Thus we have explained the *sense* in which the Maxwell-satisfying vector Bessel waves satisfy the paraxial wave equation: all that is needed is that f_{BW} be expanded to $O(\theta_k^2)$ and no higher.

3.2 Numerical Demonstrations of Mode Mixing

3.2.1 Parameters and Dielectric Stack Designs

Unless otherwise stated, the results in the rest of this chapter are for a spherical M2 of radius $R = 100\,\mu m$, with $z_1 = 0$ and $z_d = 10\,\mu m$ (see Fig. 2.2). (The absolute units of course, do not matter, microns could be replaced everywhere with meters. Since optical applications are in mind, however, we will use a scale suited for visible and NIR light.) The different mirrors used for M1 will be denoted MIA, MIB, MII, MIII, MIV. MIA is a perfect conductor, with $r_p = r_s = -1$. MIB has $r_p = r_s = +1$, corresponding to a theoretical perfect magnetic conductor, or alternatively corresponding to the following "virtual mirror" situation: if a reflection-symmetric metallic cavity is formed by reflecting the M2 dome about $z = 0$, the modes of this cavity with odd parity in E_T, are, in the upper half-plane, identical to the modes of the dome cavity with M1 = MIA. (The even parity modes are the M1=MIB modes of the dome cavity.)

MII and MIII are dielectric mirrors made with the standard high-reflection design of alternating layers of high and low refractive index, which each layer having an optical thickness of $\lambda_d/4$ (a physical thickness of $\lambda_d/(4n)$, where n is the refractive index). This quarter-wave design optimizes the reflection of light of wavelength λ_d at normal incidence. MII is the standard commercial design with the high index material at layer 1 (the surface layer). MIII is the alternate (and less reflective) design with the low index material being at layer 1. The indices of refraction are $n_{lo} = 3.0$ and $n_{hi} = 3.5$, corresponding to AlAs and $Al_{1-x}Ga_xAs$ for a certain mixing ratio x.[8] MIV is the same as MII except that layer 1 has an optical thickness of $1\lambda_d$ instead of the usual quarter-wave thickness. The design MIV, with an additional thin GaAs quantum well in the center of layer 1, is the design used by Michael Raymer [15]. The reflection behavior of MIV is quite similar to the reflection behavior for MIII (not MII). The tables in Appendix C give some reflection data for MII and MIII, although Sec. 3.3 is necessary to understand the appendix. For the characterization of stacks, we will frequently use the parameter $k_d \equiv 2\pi/\lambda_d$.

---

[8]Some of the results here are obtained with the unnecessarily "precise" values $n_{lo} = 3.003$ and $n_{hi} = 3.51695$

Figure 3.3: Instantaneous vector LG mode patterns A and B. All of the vectors in A rotate counterclockwise and all of the vectors in B rotate clockwise. (Rotating the page ccw or cw advances the fields A or B correctly.)
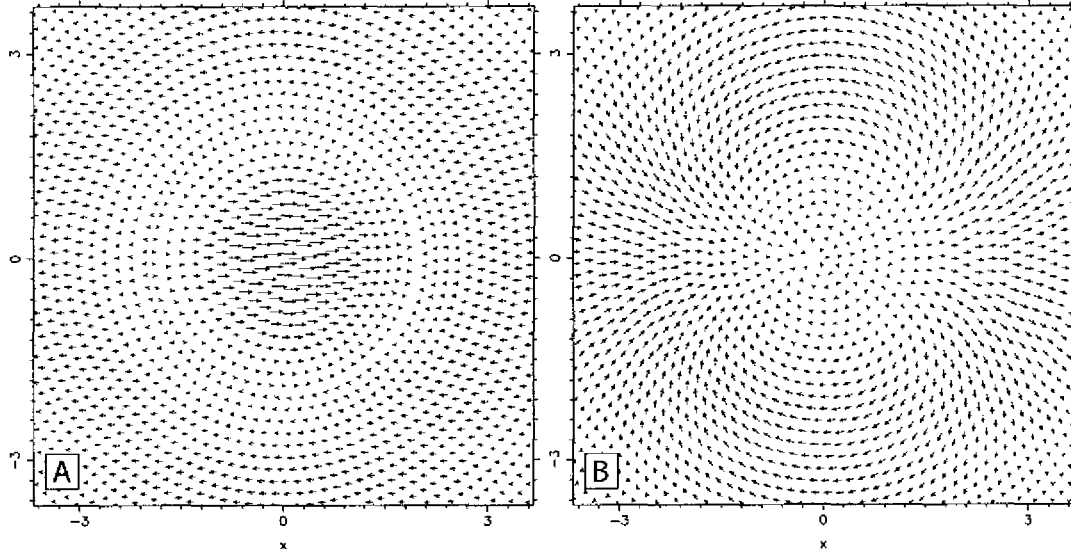
3.2.2 A M1 Mirror of Constant Phase Shift: $\alpha = 0$

Figure 3.3 shows a $t = 0$ snapshot of the vector LG modes $A \equiv LG_1^0 \sigma^+$ and $B \equiv LG_0^2 \sigma^-$ (see Table 3.2). The vector fields are plotted using the $LG_p^l(\rho, \phi)$ expression (3.14) with $\lambda = 790.287\,\text{nm}$, $h = 0.0915708$ ($w_0 = 2.74712\,\mu\text{m}$, $z_R = 30\,\mu\text{m}$). Modes A and B have total angular momentum $m = 1$ and transverse order $N = 2$. Mode A has spin $s = +1$ and orbital angular momentum $l = 0$, while mode B has spin $s = -1$ and orbital angular momentum $l = 2$. Since mode A has $p = 1$, there is one nodal ring at finite nonzero radius, as seen clearly in the figure; mode B has $p = 0$ and therefore has no such nodal rings (although it clearly has an antinodal ring). Because modes A and B are both written as the product of a scalar field and a polarization vector (rather than a sum of such terms), the polarizations of A and B are both uniform in the sense that they are circular everywhere (although the phase varies in space). The time evolution of the fields can be seen by rotating the printed figures about their centers. A superposition of A or B and its exactly degenerate counterpart freezes the unsigned direction of the vectors in the plots. Here the polarization is everywhere linear but the linear direction changes with ϕ.

Before calculating with a dielectric stack, we first consider a conducting cavity (M1=MIA) with length $z_d = 10\,\mu\text{m}$, $R = 100\,\mu\text{m}$, and $z_e = 6\,\mu\text{m}$. The $N = 2$

Figure 3.4: E_T for numerically calculated modes of a conducting cavity. The cross-sections are plotted at $z \approx \lambda/4$. The inset shows the E_x field in the x-z plane. The units are μm.
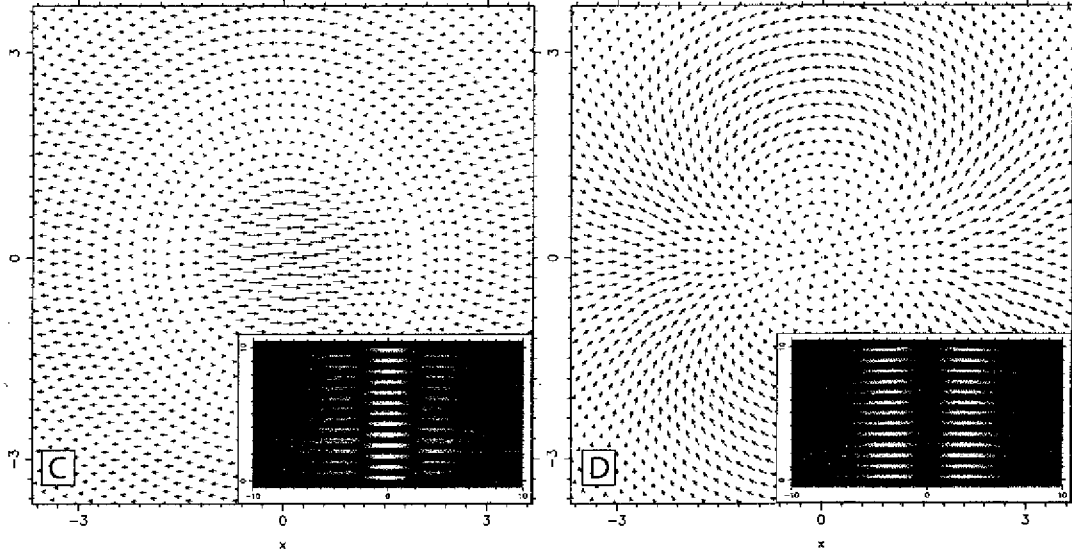
family of Gaussian beams with parameters ($\lambda = 790.287\,\mathrm{nm}, h = 0.0915708, w_0 = 2.74712\,\mu\mathrm{m}, z_\mathrm{R} = 30\,\mu\mathrm{m}, k = 7.95051\,\mu\mathrm{m}^{-1}, \nu = 3.79346 \times 10^{14}\mathrm{Hz}$) fits into this cavity according to paraxial theory (Eqs. (3.11, 3.7)). The numerical results using variant 1 of the two basis method yields two modes, C and D, which are displayed in Fig. 3.4. Here $k_C = 7.950539\,\mu\mathrm{m}^{-1}$ and $k_D = 7.950570\,\mu\mathrm{m}^{-1}$, in correspondence with the paraxial prediction. The frequency splitting is $\Delta\nu \equiv \nu_D - \nu_C = 1.5\,\mathrm{GHz}$. For these calculations, $\Delta_\mathrm{n} \approx 1 \times 10^{-7}$.

Comparing between Figures 3.3 and 3.4, one can see little difference. To obtain a mode mixing angle, we must take inner products between two vector fields. Our calculated inner products $\langle X, Y \rangle$ are simply a sum over the complex dot products of corresponding E_T field vectors, using the same discrete locations shown in the plots

$$\langle X, Y \rangle = \sum_j \boldsymbol{E}_{\mathrm{T}X}(\rho_j, \phi_j) \cdot \boldsymbol{E}^*_{\mathrm{T}Y}(\rho_j, \phi_j). \tag{3.36}$$

We expect

$$A \odot B = 0,$$
$$C \odot D = 0,$$
$$(e^{-i\varphi_C}C) \odot A \approx (e^{-i\varphi_D}D) \odot B \approx \cos\alpha,$$
$$-(e^{-i\varphi_C}C) \odot B \approx [(e^{-i\varphi_D}D) \odot A]^* \approx e^{i\beta}\sin\alpha, \tag{3.37}$$

where the product denoted by $\odot$ is defined to be the normalized version of the inner product of the fields,

$$A \odot B \equiv \frac{\langle A, B\rangle}{\sqrt{\langle A, A\rangle\langle B, B\rangle}}, \tag{3.38}$$

and where the phase correction factors of C and D are defined by

$$\varphi_C = \arg(C \odot A),$$
$$\varphi_D = \arg(D \odot B). \tag{3.39}$$

In many cases, including the one following, the calculated φ_C and φ_D are very small and we simply report the uncorrected $\odot$ products.

The numerical calculation yields

$$
\begin{array}{ll}
A \odot B = 2 \times 10^{-12} & C \odot D = 0.0061, \\
C \odot A = 0.99996 & D \odot B = 0.99990, \\
-C \odot B = 0.0074 & D \odot A = 0.0135, \tag{3.40}
\end{array}
$$

with the imaginary parts of each $\odot$ product having a magnitude smaller than 10^{-5}. While this data possibly indicates a positive mixing angle around 0.01 radians, our interpretation is that this is due to numerical error and that $\alpha = 0$. Other calculations with MIA and MIB mirrors indicate similarly small values of α, and the likely case is that the modes of such cavities are indeed pure LGσ modes. One of the most important result from this calculation is that the data supports the framework of the theory: the two new mode patterns are essentially orthogonal as hypothesized.

These data also support the result of a conducting microwave cavity investigation done in the late 1970's by Erickson and Cullen [36, 49, 44], and the mid 1980's by Yu and Luk [45, 46, 47]. This work, which does not address the possibility of mixing,

Figure 3.5: Scan of Δ_r vs. $(z_d \operatorname{Re} k)$ indicating the presence of the near degenerate modes shown in Figs. 3.6 and 3.7.
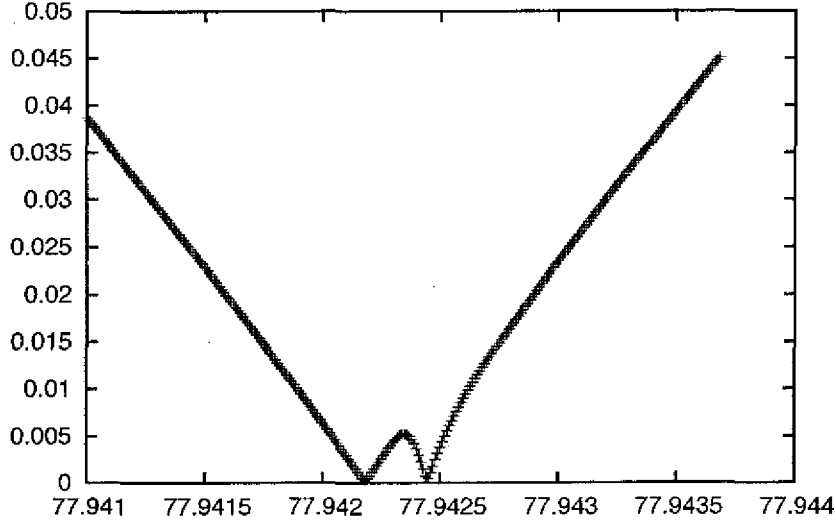
shows experimentally that the eigenmodes of a conducting Fabry-Perot cavities have pure LGσ mode patterns, and derives theoretically the frequency splitting of these modes. The final result is a sum of three different corrections found using three different perturbation theories. In Appendix D we give a fundamental result for electromagnetic perturbation theory that is used in one of the correction calculations. It appears that that this body of work—not only the authors but also the results—has been forgotten by many working with Fabry-Perot cavities (including VCSELs) today. The result of the perturbation theory will be given in Eqs. (3.47, 3.48). Eq. (3.47) predicts $k_A = 7.950538 \mu m^{-1}$, $k_B = 7.950570 \mu m^{-1}$, in very close agreement with our numerical k_C and k_D, given earlier in this section.

3.2.3 A M1 Bragg Mirror: $\alpha \neq 0$

Next we consider M1=MIV for modes with $\operatorname{Re} k \approx k_d = 7.95051 \, \mu m^{-1}$. (Here $\lambda_d = 790.287 \, nm$.) Since the design MIV has $\phi_s \approx \phi_p \approx 0$ (not π) near $\theta_k = 0$, we use Eq. (3.11) with the RHS equal to 24.5π and $k = 7.95051 \, \mu m^{-1}$ to calculate $z_d = 9.801$. Using this value of z_d and $N = 82$ layers with $n_{hi} = 3.5$ and $n_{lo} = 3$, we find the actual modes at $k_C = 7.952472 - i2.67 \times 10^{-7} \, \mu m^{-1}$ and $k_D = 7.952499 - i2.61 \times 10^{-7} \, \mu m^{-1}$. Figure 3.5 shows a scan of Δ_r vs. $(z_d \operatorname{Re} k)$ with the two minima indicating modes C and D. As for the previous calculations, $\Delta_n \approx 1 \times 10^{-7}$. The normalized inner

Figure 3.6: Snapshot of E_T for C and D for M1=MIV. Here $\alpha = 21.1°$.
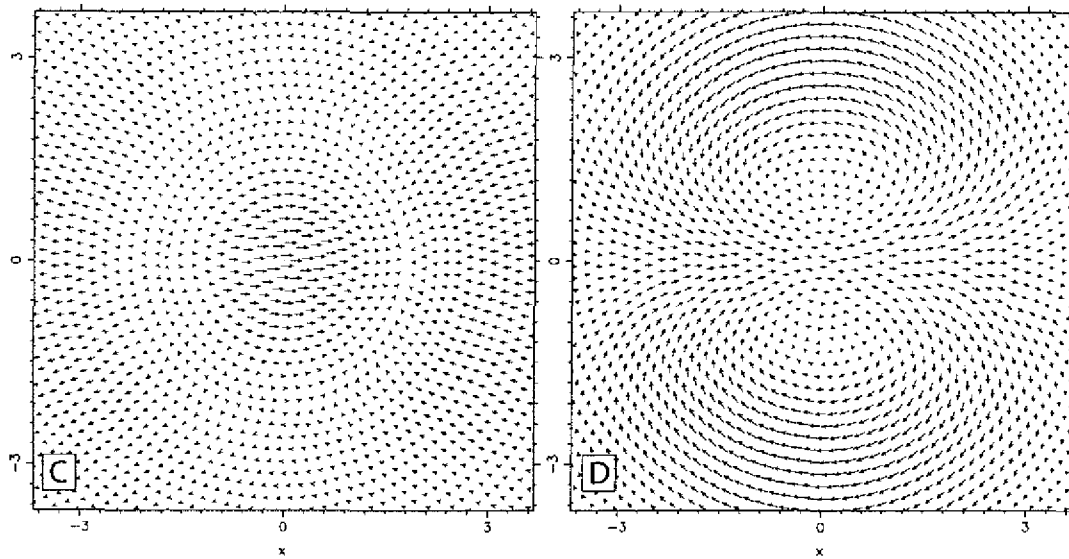
products and the corresponding predictions of α are

$$C \odot D = -0.0009 - 0.0002i,$$

$$C \odot A = 0.9325 + 2 \times 10^{-6}i \qquad \Rightarrow \alpha = 21.17°,$$

$$D \odot B = 0.9330 - 0.0003i \qquad \Rightarrow \alpha = 21.09°,$$

$$-C \odot B = 0.3605 + 0.0001i \qquad \Rightarrow \alpha = 21.13°,$$

$$D \odot A = 0.3597 + 2 \times 10^{-5}i \qquad \Rightarrow \alpha = 21.08°. \qquad (3.41)$$

Here we have demonstrated a well-defined nonzero mixing angle $\alpha \approx 21.1°$ for modes well into the paraxial limit. Again it appears that $\beta = 0$.

The mixed wave functions are shown in Figs. 3.6 and 3.7. Figure 3.6 shows a snapshot of E_T while Fig. 3.7 shows the polarization ellipses. Because the radial dependence of the two constituent mode patterns A and B are different, the result of mixing produces modes for which the polarization varies with ρ as well as ϕ. Ref. [26] contains a figure of the polarization fields for two modes which have nearly maximal mixing ($\alpha = 39°$). The variation of amplitude and polarization with ρ is very pronounced for these mode patterns.

To check these results we have redone the entire calculation using the Bessel wave method (disallowing $\theta_k > 0.6$ to make the solution stable). The resulting four values

Figure 3.7: Polarization fields for C and D. The tick inside each ellipse shows the instantaneous phase at $t = 0$. The hollow ellipses indicate ccw rotation and the filled ellipses indicate cw rotation. The polarization is seen to vary in ρ while the phase varies in ϕ.
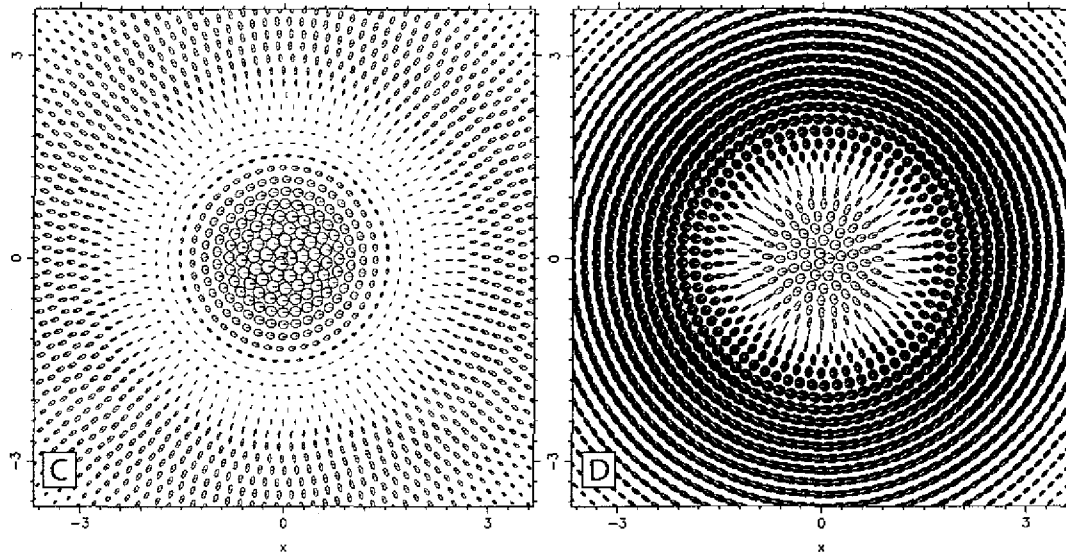
of α (in the order used in Eq. (3.41)) are 23.48°, 23.29°, 23.44°, and 23.27°, with Δ_{n} being 3.6×10^{-5} for mode C and 1.9×10^{-5} for mode D. The results differ from those of Variant 1, but the agreement is sufficient to demonstrate that the mixing angle is not 0. An attempt was made to redo the calculation with Variant 2. However, results seemed to be unstable, depending significantly on the nature of the seed equation.

Parenthetically, we can check numerical calculation of $\operatorname{Im} k$. A stack calculation gives $r = 0.99999487$ for the field reflectivity of the Bragg mirror at normal incidence when the incident wave number is $k = \operatorname{Re} k_C$. Using Eq. (1.25), this predicts a Q value of 1.520×10^7. The numerically found Q value, using (1.24) with k_C, is 1.491×10^7. The agreement between the two Q values indicates that, even though $|\operatorname{Im} k|$ is smaller than $10^{-7}|k|$, its value has been predicted to within 2 percent. (The agreement may even improve if the finite range of the θ_k for the mode is taken into account.)

3.2.4 A Further Test

Regarding the results of the previous section, one can still raise the objection, that, because the difference in k values for modes C and D is so small, the difference in the calculated modes may be a numerical artifact. This is a difficult objection to

put to rest, as the modes are near degenerate except when the mode geometry is extremely nonparaxial, in which case the mode cross sections are very different from superpositions of LG mode patterns. The best that can be done is to look at a cavity shape where the modes are marginally paraxial. Here we show a case in which the separation in $\operatorname{Re} k$ is 0.001, only 10 times smaller than the local average mode spacing (numerically determined by finding (at least) 11 minima of Δ_n between wavenumbers 8.0 and 8.1). Using a MIV mirror ($N = 82, k_d = 8.0000\mu\text{m}^{-1}, \lambda_d = 0.7854\mu\text{m}$), the parameters for this example are ($z_d = 10\,\mu\text{m}, R = 11\,\mu\text{m}, z_e = 0$) with the results ($\alpha = \{40.2°, 40.5°, 36.6°, 36.4°\}$, $C \odot D = 0.034$, $k_C = 8.064977 - i2.4 \times 10^{-7}\,\mu\text{m}^{-1}$, $k_D = 8.065989 - i1.6 \times 10^{-7}\,\mu\text{m}^{-1}, \Delta_n < 3 \times 10^{-9}$) where the α values are in the usual order and the value $w_0 = 0.71\,\mu\text{m}$ has been used to generate A and B via (3.14). (With the average wavelength of C and D, this w_0 corresponds to $h = 0.35$.) The large (but not maximal) mixing angle in this example with large $k_D - k_C$ demonstrates that general mixing angles are not likely to be a numerical error, caused by some difficulty in resolving nearly degenerate modes.

3.3 Calculated Behavior of the Mixing Angle

3.3.1 Mixing Curves

The previous section has shown one example of a mirror of constant phase shift, for which there was no λ_d, one example where $\lambda_C \approx \lambda_D \approx \lambda_d$, and one example where $\lambda_C \approx \lambda_D$ are *not* approximately equal to λ_d. In this section we investigate the behavior of the mixing angle as

$$\Delta k_r \equiv (\operatorname{Re} k - k_d)/k_d = (\lambda_d - \lambda)/\lambda \tag{3.42}$$

is varied. Δk_r can be varied by changing z_d, R, or λ_d. While these three parameters affect the modes in different ways, we find that, over a certain small range, the variation of any of these parameters changes the mixing angle α significantly. In general, the mode parameters other than α, such as w_0, k, and $h(w_0, k)$, should only suffer a slow, linear change. We choose to change λ_d in order to vary Δk_r and, consequently, α.

Figure 3.8 shows the variation of α vs. Δk_r for three different cavity designs: C1, C2 and C3. Appendix C gives details. Our data is obtained by using Variant 1 of

Figure 3.8: Mixing angle behavior for three cavities. The k_0 values are given in microns. The fitted curves have an arctan shape with two adjusted parameters. Labeled data points give $\operatorname{Re} k$ in terms of the value of k_0 for that data series.
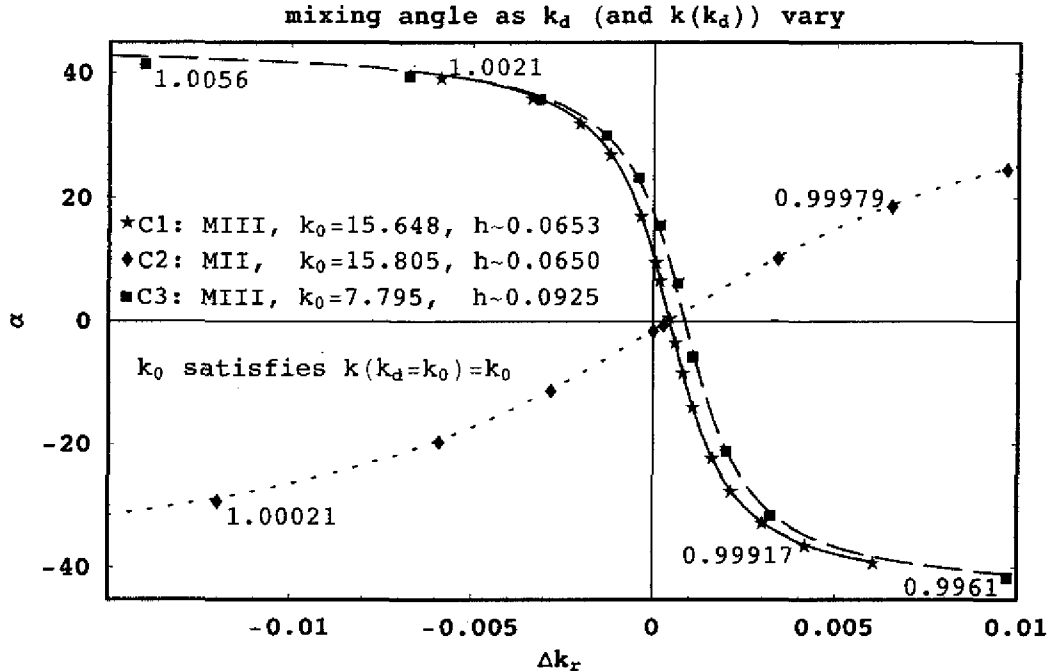
two-basis method. The mixing angle is seen to be asymptotically maximal for large $|\Delta k_r|$ and to sweep through 0 in a relatively narrow region. Each curve has been fit with a two-parameter function:

$$y = \frac{1}{2} \arctan[\zeta(x - \xi)]. \tag{3.43}$$

To vary Δk_r, z_d and R were kept fixed while λ_d was varied (this of course caused λ to vary as well). Each mixing angle datum shown in the figure is the mean of the four α calculated as before. The inner products used to obtain the mixing angles in the figure are between the cross section of the mode C (or D) and an actual cross section of a mode calculated with MIA or MIB, (determined by the phase shift the Bragg mirror approached at $\theta_k = 0$, $k = k_d$). These latter mode patterns correspond very well with the analytically calculated A and B. There are several other difference in the way these values have been calculated from those in the previous demonstration.

61

Figure 3.9: Anticrossing of the modes C and D for C1. Stars: frequency splitting (left axis); Diamonds: mixing angle (right axis).
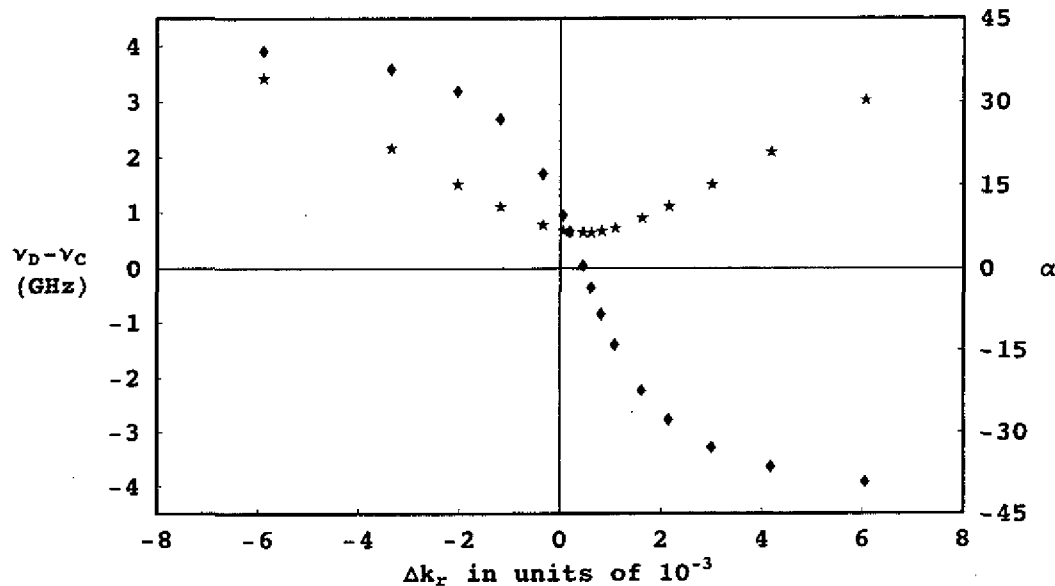
However, re-examination of several individual data points have shown very good agreement between these data and the results of the more careful calculation.

Cavity C1 and C3 both have M1=MIII while C2 has M1=MII. Each cavity has $R = 100\,\mu m$, $z_d = 10\,\mu m$, $z_1 = 0$, $z_e = 0$, $n_{hi} = 3.51695$, $n_{lo} = 3.003$, and $N = 72$ stack layers. For each cavity we can assign a special value k_0 (and λ_0) where the average $\mathrm{Re}\,k$ for the modes C and D is equal to k_d. These values of k_0 are listed in the figure; for C1 and C2 the wavelengths involved are around 400 nm, while for C3 λ is around 800 nm. To show the extent over which k and k_d vary, several data points have been labeled with their $\mathrm{Re}\,k$ values in terms of k_0. The value of k_d at these points is then given by $k_d = \mathrm{Re}\,k/(1 + \Delta k_r)$. It is seen that as k_d is changed from k_0, the real part of k, also starting at k_0, follows behind k_d.

As the mixing angle varies over 90° as Δk_r is scanned, the patterns of modes C and D are exchanged, as can be seen theoretically from the transformation (3.16). This change of character is associated with an anticrossing of modes C and D. Figure 3.9 shows the hyperbolic shape of $\nu_D - \nu_C$ for C1 which characterizes the anticrossing.

The minima of the frequency difference occurs where $\alpha = 0$, as is expected for an anticrossing.

3.3.2 The Role of Birefringence

In these simulations we have given the Bragg mirrors a large number of layers in order to lay aside any effects that come from finite resonance width. Accordingly we can describe the mirror reflectivities r_s and r_p by $e^{i\phi_s(k,\theta_k)}$ and $e^{i\phi_s(k,\theta_k)}$. Since we are in the paraxial limit, we can expand the plane wave reflection phase functions for a given stack design[9] as

$$\phi_s(\Delta k_r, \theta_k) = \phi_0(\Delta k_r) + \epsilon_s(\Delta k_r)\theta_k^2 + \delta_s(\Delta k_r)\theta_k^4,$$
$$\phi_p(\Delta k_r, \theta_k) = \phi_0(\Delta k_r) + \epsilon_p(\Delta k_r)\theta_k^2 + \delta_p(\Delta k_r)\theta_k^4. \tag{3.44}$$

Note that $\phi_s = \phi_p$ at $\theta_k = 0$ since at normal incidence p polarized waves becomes indistinguishable from s polarized waves. The expansion is even in θ_k because construction of the stack transfer matrices is manifestly even in $(\sin\theta_k)$. Since Δk_r is relatively small, the coefficients ϕ_0, $\epsilon_{s/p}$, and $\delta_{s/p}$ can be expanded to first or second order in Δk_r. The expansion has been introduced here to show how ϕ_s and ϕ_p depend on θ_k and Δk_r; the exact values of $r_{s/p}$ are used in the numerical calculations. The expansion (3.44) is used in Chapter 4 in the search for simplified models which describe the mixing behavior.

We have claimed that the most important property of the stack for affecting the mixing angle is the birefringence, which is characterized by the function $\phi_s - \phi_p$. Part of the numerical evidence for this claim comes from running our programs with Bragg mirrors and setting r_p to the calculated r_s (or setting r_s to the calculated r_p). These tests with zero birefringence do indeed indicate a mixing angle of 0 (or a small mixing angle at least). Table 3.3 shows the results of these tests for several points on Fig. 3.8. These results indicate that mirror birefringence is necessary to have large or intermediate mixing angles.

The natural thing to do now is to replot the mixing angle curves of Fig. 3.8 with the pertinent quantity $[\phi_s(k, k_d, \theta_{k,\text{eff.}}) - \phi_p(k, k_d, \theta_{k,\text{eff.}})]$ along the x axis. There is some choice involved in how to define the effective polar angle, $\theta_{k,\text{eff.}}$. We choose to

---

[9]The word "design" here refers to the specification of the refractive indices (taken to be k-independent) and the layer thicknesses *relative to the unspecified scale* λ_d .

cav. design	orig. mean α	α vals. for $r_\mathrm{p} \to r_\mathrm{s}$	α vals. for $r_\mathrm{s} \to r_\mathrm{p}$
C1	$-36.5°$	$0.71, 1.09, 0.17, 0.96°$	$0.76, 1.08, 0.22, 0.93°$
C2	$-29.2°$	$0.12, 0.06, -0.08, 0.25°$	$0.07, 0.10, -0.25, 0.11°$
C3	$41.4°$	$0.93, 0.79, -0.51, 0.30°$	$0.68, 0.59, -0.49, 0.27°$

Table 3.3: Mixing angles when birefringence is zero. The situations tested are selected from points on Fig. 3.8.

simply set

$$\theta_{k,\mathrm{eff.}} = [\bar{\theta}_k(h, A) + \bar{\theta}_k(h, B)]/2 \qquad (3.45)$$

where the mode symbols A and B determine the values of p and l. The values of $\theta_{k,\mathrm{eff.}}$ for C1, C2, and C3 are 0.07433, 0.07399, and 0.1055 radians, respectively. We are ignoring the fact that h changes slightly as k and k_d vary across the plot. (This variation in h, when included, makes very little difference in the plot.)

The resulting plot is shown in Fig. 3.10. Comparing these mixing curves with those of Fig. 3.8, one can see that $(\phi_\mathrm{s} - \phi_\mathrm{p})$ is a more universal indicator of α than Δk_r. It is interesting that the data series for C1 and C2 are nearly on top of each other, having fitting parameters $\zeta = -3325$ and $\zeta = -3438$ via (3.43), while the C2 curve was the odd-ball of Fig. 3.8. C1 and C2 have different M1 designs, and the coincidence of their curves indicates that $(\phi_\mathrm{s} - \phi_\mathrm{p})$ is a very good universal indicator of the mixing angle *for modes which fit the same into the given cavity shape*, that is, for cavity-mode combinations with essentially the same h, R/λ, and $(z_\mathrm{d} - z_\mathrm{l})/\lambda$, so that they look the same when outlined in a side view. Unfortunately the C3 curve has $\zeta = -1612$, differing significantly from the C1 and C2 curves. The version of this plot shown in Ref. [26], which has *nearly the same* relative ζ-disagreement between C3 and C1/C2, uses h itself as the effective polar angle. The relative insensitivity of the C3 disagreement to the definition of the effective angle may indicate that the reduction of the mode outline to a single effective Bessel wave is too simplistic.

A surprising agreement of all three curves is seen in Figure 3.11, where α is plotted against $(\epsilon_\mathrm{s} - \epsilon_\mathrm{p})$, as defined by Eq. (3.44). The arctan fits yield $\zeta_{C1} = -18.17$, $\zeta_{C2} = -18.83$, and $\zeta_{C3} = -18.53$. The reason for the apparent independence of differences in $\bar{\theta}_k$ is not fully understood, although this agrees with the result of the curious perturbation theory of Sec. 4.2. Although more series of data will be needed to support a compelling statement of an empirical function such as $\alpha \approx (1/2) \arctan[(-18.5)(\epsilon_\mathrm{s} -$

Figure 3.10: Mixing angle vs. birefringence. The plane wave reflection phases were evaluated at $\theta_k = \theta_{k,\text{eff}}$ for each cavity.
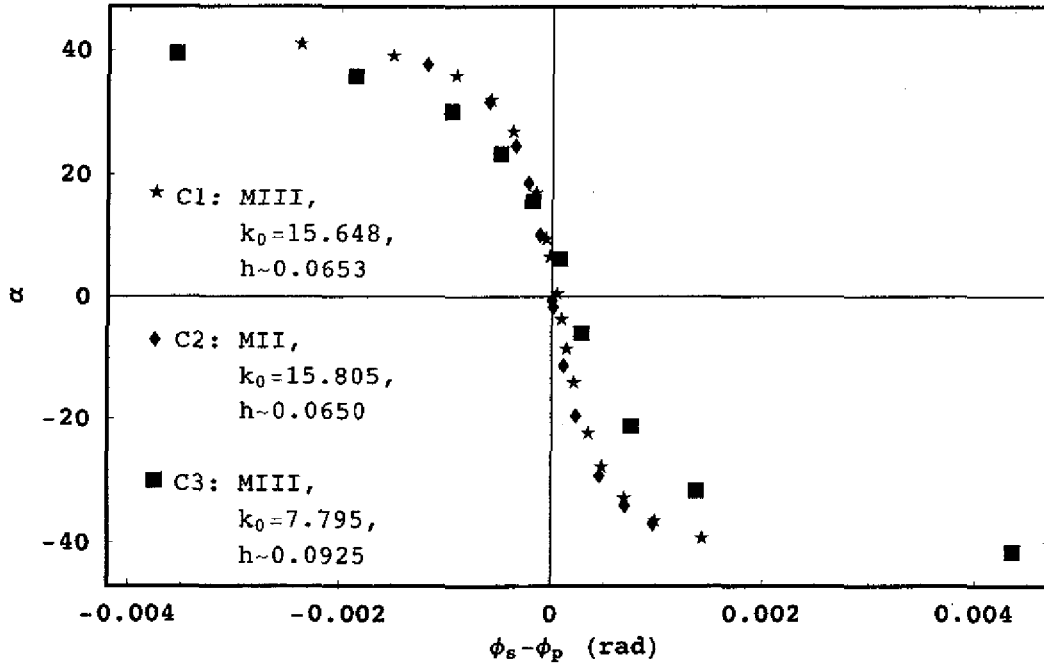
$\epsilon_p - \xi)$], we may conclude from this section that the mixing behavior is primarily governed by the birefringence of the Bragg mirror.

If one looks at the distribution of plane wave components in the the mixed mode pairs as a function of α, one sees that, at $\alpha = 0$, the distributions for s and p polarized light are nearly identical, but, near $\alpha = \pm45°$, one mode is predominantly s polarized and the other is predominantly p polarized. Thus the mixing may be loosely interpreted (at least for the $N = 2, m = 1$ mixable pair, which we have studied) as an anticrossing between an s polarized and a p polarized mode. The role of $\epsilon_s - \epsilon_p$ now has a somewhat intuitive interpretation: when the $|\phi_s - \phi_p|$ is large, the modes are selected so that one has mostly s polarization and one has mostly p polarization. As $|\phi_s - \phi_p|$ increases further, the energy splitting increases because the s mode and the p mode are seeing a greater difference in effective cavity length (due to the phase shift at M1). Figure 3.12 shows the s and p plane wave distributions for two of the C1 mode pairs. The figure shows that if the mode has two θ_k maxima, it is the maxima at larger θ_k which experiences the strong s-p selection, as this is where $|\phi_s - \phi_p|$ is

Figure 3.11: Mixing angle vs. the reflection phase expansion coefficient difference $(\epsilon_s - \epsilon_p)$. This graph essentially shows the dependence of the mixing angle on the reflected phase birefringence evaluated at any value of θ_k taken to be the same for C1, C2, and C3, regardless of fact that the distributions $g(\theta_k)$ are different. Surprisingly, the three curves have approximately the same width, indicating that the quantity $(\epsilon_s - \epsilon_p)$ itself is the most important mixing quantity.
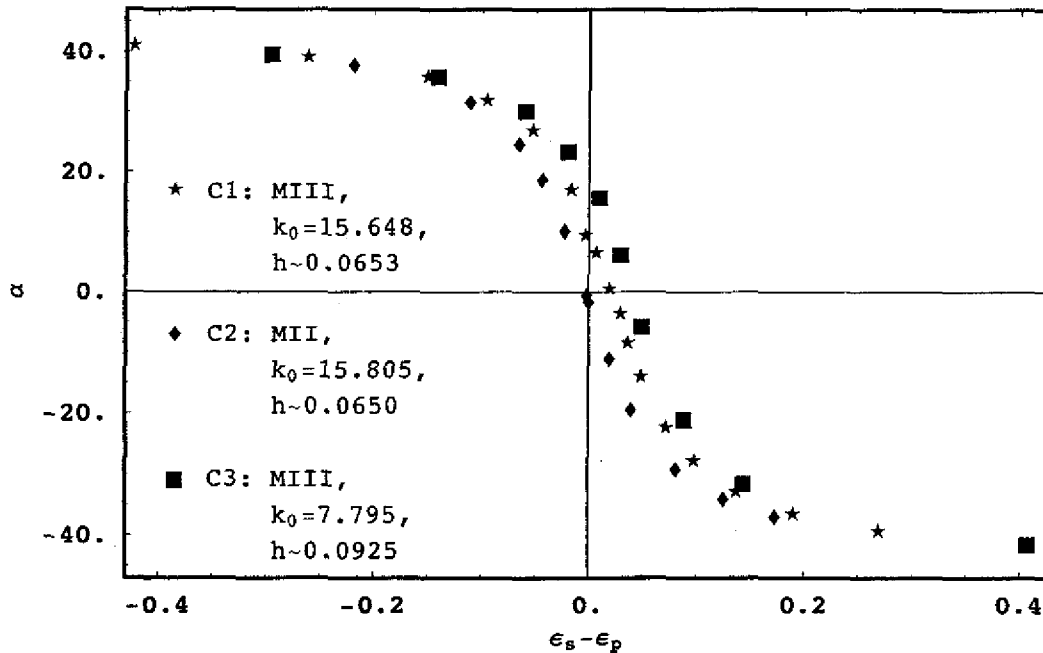
greatest. The plane wave distributions for the mode pair near $\alpha = 0$ are seen to agree well with the pure LG mode distributions shown in Fig. 3.2.

3.4 Outlook for Spin-Orbit Coupling

3.4.1 Criteria for Observation

In order to observe spin-orbit coupling, one must be able to resolve the paraxial degeneracy. This means that the spectral width of the cavity resonator, given by $\gamma_k \equiv -2\text{Im}\,k$, must be less than the splitting, $\delta_k \equiv k_D - k_C$. (This inequality, $\gamma_k < \delta_k$, is the Rayleigh criterion for resolvability; in order to make practical use of mode mixing, one must have $\gamma_k \ll \delta_k$.) Estimation for the spectral width is straightforward
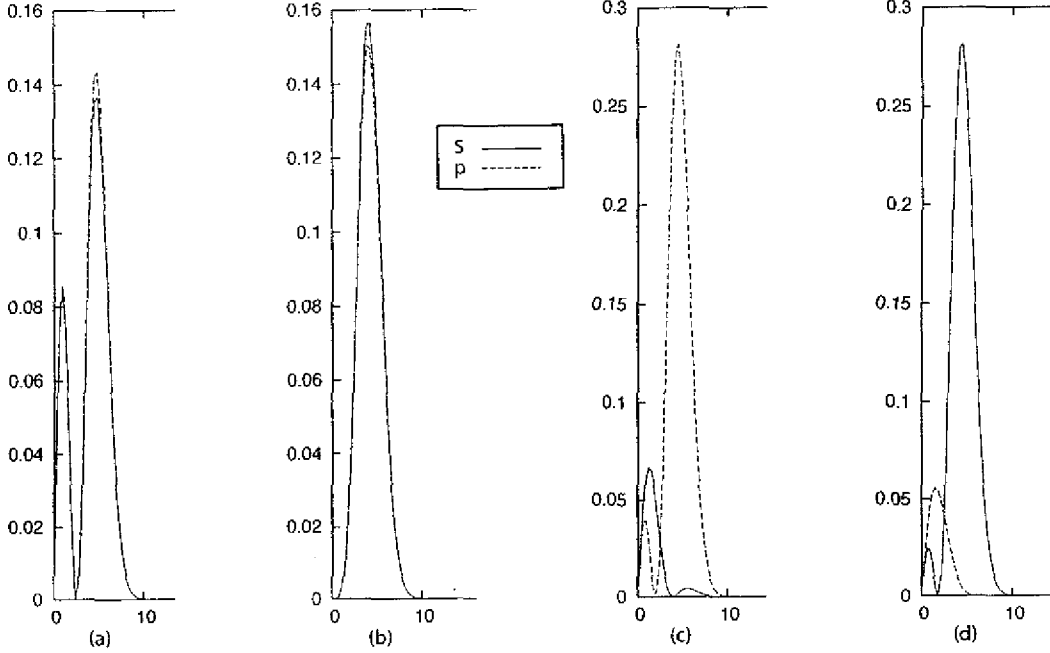
Figure 3.12: Plane wave distributions $\sin\theta_k|S_d(\theta_k)|^2$ and $\sin\theta_k|P_d(\theta_k)|^2$ for two mode pairs from the C1 data. The x axis is θ_k in degrees. Plots (a) and (b) are the C and D modes of the C1 data point having $\alpha = 0.5°$. Plots (c) and (d) are the C and D modes of the C1 data point having $\alpha = 39.1°$. It is clear that the modes are predominantly s or p polarized for maximal mixing. Since maximal mixing is the generic situation, this suggests that paraxial mode mixing can be interpreted as an anticrossing between an "S mode" and a "P mode".

via Eq. (1.25). Defining the (geometric) mean power reflectivity $R_{12} \equiv (R_1 R_2)^{1/2} = |r_1 r_2|$, the width is given by

$$\gamma_k = \frac{1 - R_{12}}{L\sqrt{R_{12}}}. \tag{3.46}$$

Clearly, higher mirror reflectivities produce better resolvability. The auspicious fact that γ_k is inversely proportional to L (and longer resonators are simple to build) is unfortunately eclipsed by the observation, which we will show next, that $\delta_k \propto 1/L^2$ when $L^2 \gg z_R^2$.

As can be seen from the frequency splitting plot in Fig. 3.9, δ_k increases away from the transition region of α. Furthermore, the data given in Appendix C show that the relative increase of $|\gamma_k|$ (as $|\alpha|$ increases) is slower than the relative increase of δ_k, at least over the region studied. This is good news for the observability of $|\alpha| \approx 45°$.

Maximally mixed modes should become observable before transitional modes.

To get an idea of what type of cavities have the largest δ_k, we turn our attention to conducting cavities (M1 = MIA) for which δ_k has been analytically determined via a perturbation theory [47, 46] which is discussed in Appendix D. Fortunately, our numerical data have indicated that the δ_k for conducting cavities is close to (usually within 50% of) the minimum δ_k for the cavities of the same shape but with a dielectric M1 mirror. Thus δ_k for conducting cavities can be used to indicate the observability of the transition region of α, and one has the happy knowledge that, in the asymptotic region of α, δ_k will be somewhat larger. (However, as $|\Delta k_r|$ increases, $|\gamma_k|$ will eventually start rapidly increasing as the mode frequency falls off the stopband of the mirror. This limits how much the splitting observability can be improved by increasing $|\Delta k_r|$.)

The result at the end of Ref. [47], for the $LG_p^l \sigma^s$ eigenmode of a plano-concave conducting cavity, is that the perturbation in frequency, $\Delta \nu$, from the frequency that would be expected from paraxial theory (3.12), is given by

$$\Delta \nu = \frac{c}{16\pi k L R}(2p^2 + 2p|l| - l^2 + 2p + |l| - 2 \pm 4|l|), \qquad (3.47)$$

where L is the cavity length and the $+$ sign is chosen if $|l + s| < |l - s|$ and the $-$ sign is chosen if $|l + s| > |l - s|$. Using this formula with the modes associated with mode patterns A and B, we obtain

$$\delta_{k,\text{cond}} = k_B - k_A = \frac{1}{4kLR}. \qquad (3.48)$$

This relation can be transformed, via $R = L + z_R^2/L$ and other paraxial relations, into

$$\begin{aligned}
\delta_{k,\text{cond}} &= \frac{k}{4} \frac{1}{k^2 L^2 + k^2 z_R^2} \\
&= \frac{k}{4} \frac{1}{k^2 L^2 + \frac{4}{h^4}} \\
&= \frac{k}{4} \frac{1}{k^2 L^2 + k^4 w_0^4/4}. \qquad (3.49)
\end{aligned}$$

For our data for C1, C2, and C3, we have $z_R^2 \approx 9L^2$, so that we are entering the

$z_R^2 \gg L^2$ regime where

$$\delta_{k,cond} \approx k\left(\frac{h}{2}\right)^4. \tag{3.50}$$

A simple calculation shows that the transmittance $T_{12} = 1 - R_{12}$ necessary to beat the Rayleigh criterion ($\delta_k > \gamma_k$) is given by

$$T_{12} < kL\left(\frac{h}{2}\right)^4 = \frac{kL}{(kw_0)^4}, \tag{3.51}$$

where the approximation $\sqrt{R_{12}} \approx 1$ has been used.

In the other regime where $L^2 \gg z_R^2$, the condition necessary to beat the Rayleigh criterion is

$$T_{12} < \frac{1}{4kL} \tag{3.52}$$

where again we have used $\sqrt{R_{12}} \approx 1$. For reference, the quantity $1/(4kL)$ has a value of 0.0032 at $\lambda = 800$nm, $L = 10\mu m$. The primary requirement for observability in this regime is that the cavity be as short as possible. For several values of the pair (p, l), the breaking of the transverse mode degeneracy for the mode pairs ($LG_p^l\sigma^+$, $LG_p^l\sigma^-$), (these are not the mixable pairs, but they have comparable splittings), in a conducting microwave cavity has been observed by Yu and Luk [45], who have found the splittings agree well with their perturbation theory.

For a cavity filled with refractive index n, it appears that the general criterion on T_{12} is

$$T_{12} < \frac{knL}{(2knL)^2 + (knw_0)^4}. \tag{3.53}$$

In practice, k is set by the desired application, and R_{12} set by the highest mirror reflectivities reasonably achievable. One can then maximize the "observability", $\delta_{k,cond}/\gamma_k$, given by

$$\frac{\delta_{k,cond}}{\gamma_k} = \frac{1}{4kT_{12}} \frac{1}{L + \frac{z_R^2}{L}} \tag{3.54}$$

For a fixed z_R (or a fixed w_0), the observability achieves its maximum value, $1/(8kT_{12}z_R)$, at $L = z_R$. Note that when $L < z_R$, the observability actually increases

with L, in agreement with (3.51).

It is important to realize that the property of spin-orbit coupling is not confined to hemispherical cavities with one metallic and one dielectric mirror. A cavity with one or two dielectric mirrors, whether plano-concave or double-concave, should exhibit mixed modes. The reason we can make this general statement is that the discussion of Sec. 3.1.3 makes no assumptions about the mirrors. This section in fact shows that mixing will generally occur; we then numerically find this mixing in the dome cavity, which we have chosen because we have especially efficient methods for doing the numerics for this shape.[10] We note that while mixing should occur generally, we do not know that the arctan-shaped curve of the mixing *behavior* is general. The smallest tunable cavity that the author knows of is a 60 μm dome cavity with two dielectric mirrors built by Michael Raymer's group an the University of Oregon.[11] This cavity is in the $L^2 \gg z_R^2$ regime and may already be able to resolve the transverse mode splitting in the asymptotic regions of α.

Additionally, there is the possibility that mixing could occur in VCSELs. Typical VCSELs are quantum well lasers with planar-planar cavities, usually 1/2 wavelength long, with the mode being confined in the transverse directions by a circular aperture which is either cut into an insulating oxide layer or formed by etched sidewalls (as opposed to transverse confinement due to the geometric effect of a curved mirror). Because the aperture restricts the waist size, and therefore creates a beam spread by diffraction, it seems likely that VCSELs will act much like curved-mirror cavities. Regarding this subject the author is currently corresponding with Martin Achten-hagen, an author of Ref. [21] and Pierluigi Debernardi, an author of Refs. [20, 50], who have performed numerical calculations and experiments regarding the nonuni-form polarization of VCSEL modes. An interesting hybrid between VCSELs and dome cavities has been constructed by Sarangan and Peake [51], who use a shadow mask to deposit a dielectric top mirror that decreases in thickness with increasing ρ, thus making a front-surface-to-back-surface transition from planar dielectric inter-faces to curved dielectric interfaces. Overall, we believe the outlook for observation of spin-orbit coupling is quite good.

---

[10]In chronological order, the cavity methods were developed initially to study non-paraxial modes, then the reasoning of Sec. 3.1.3 was developed, and finally the mixing of LG modes was discovered after directly searching for it.

[11]The substrate for the curved mirror was made using a technique suggested by the student shop guru, David Sankovich, who apparently is in some sort of genius protection program.

Figure 3.13: Polarization fields for the $\alpha = 39.1°$ modes described in the text. See caption of Fig. 3.7.
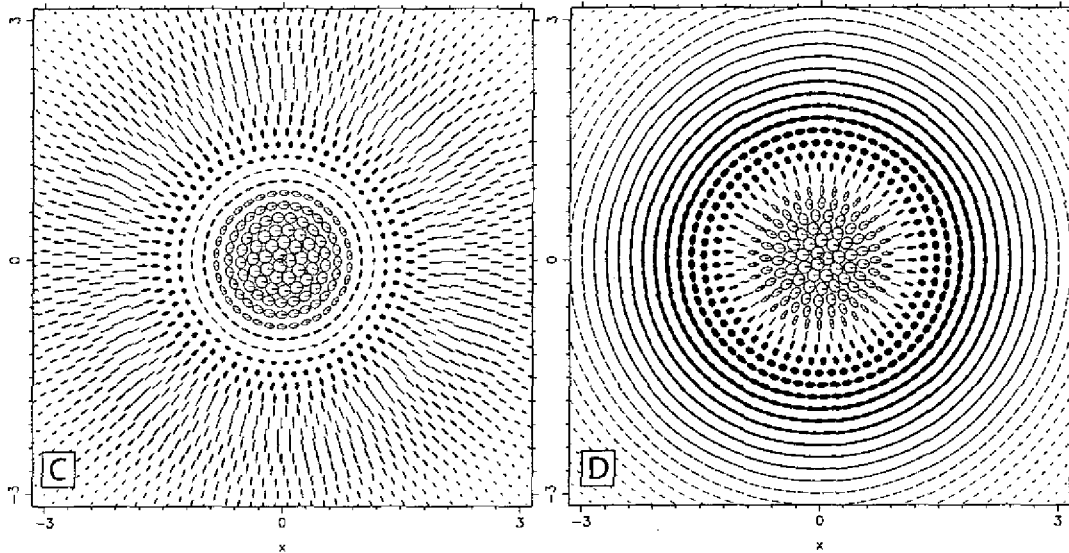
3.4.2 Possible Applications

Figures 3.13, 3.14 show the nearly maximally mixed ($\alpha = 39.1°$) modes C and D corresponding to the leftmost star point on Fig. 3.8 and the second data point listed in the C1 table of Appendix C. Mode C has circular polarization near the axis but radial linear polarization within its second radial antinode. Mode D has a transition from circular polarization in the center to azimuthal radial polarization on the outside. At intermediate radius, both modes have a low amplitude transition region in which the sense of electric field rotation of reverses. Thus the mixed modes possess a rich polarization pattern with radial dependence. The polarization patterns shown in Fig. 3.14 are frozen in place (up to the sign of the vector direction) when equal-weight superpositions of $+m$ and $-m$ is excited. These mode patterns possess the additional structure of intensity variation in ϕ. The rich polarization structure suggests that these modes may be useful as optical tweezers, which are used to manipulate molecules and other nanoparticles for a number of research and engineering applications [52, 53]. Already, modes and beams containing orbital angular momentum, and various superpositions of them, have been demonstrated or theoretically shown to trap and apply torques to nanosize particles. While superpositions of LG modes, which possess many more degrees of freedom than our single parameter α, can be made in various ways, the cavities described in our work have the possibility of being one of the

Figure 3.14: A $t = 0$ snapshot of $\boldsymbol{E}_{\mathrm{T}}$ for the $\alpha = 39.1°$ modes described in the text.
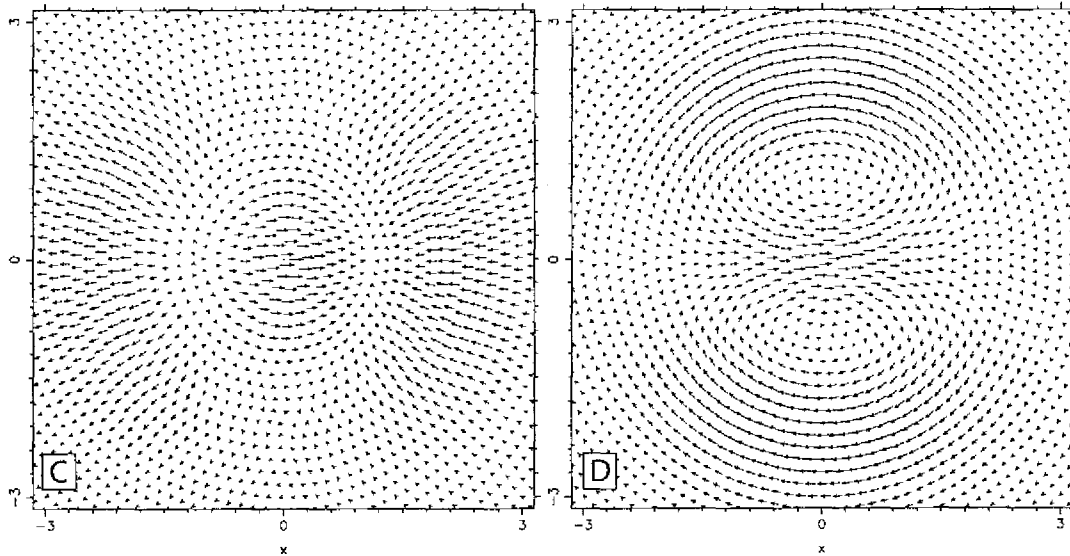
simplest possible sources of these modes and beams. Another relatively simple (and common) source of LG superposition is a computer generated, etched hologram. The cavity source in principle has the advantage of possessing a tunable parameter α, which could be useful for applications in which the concomitant frequency variations of several tenths of a percent do not matter.

Perhaps the single most useful feature regarding the cavities presented here is the fact the modes are nearly degenerate. As one example, consider the following situation. A cavity of length 100 - 200 μm is constructed with sufficiently highly reflective mirrors so that the three nearly degenerate $N = 2$ modes C, D and $E \equiv \mathrm{LG}_1^3\sigma^-$ are well resolved, but have frequencies within, say, 1 GHz of each other. The cavity parameters are fixed so that α is fixed at some value. Suppose there are embedded in the surface of M1 two nearly identical, spatially separated, polarization-sensitive, interface fluctuation (excitonic) quantum dots, QD1 and QD2. The intrinsic width of the quantum dot single-exciton states is around 2-4 GHz, and we have assumed that QD1 and QD2 are nearly identical so that the energy levels are essentially degenerate from one dot to another. Assume the cavity can be excited with sufficiently monochromatic laser pulses at the three frequencies ν_C, ν_D, and ν_E. Due to the different polarization (and intensity) patterns of modes C, D, and E, each of the pulses affects each of the quantum dots in a different way (perhaps in some cases not at all). Furthermore, the quantum dots are coupled to each other through the cavity modes.

Perhaps mode C excites (drives) QD1 but not QD2, mode D excites QD2 but not QD1, and both QD1 and QD2 couple to cavity mode E, and hence to each other. The coupling of QD1 and QD2 could perhaps be modified by pulses at frequency ν_E. This situation demonstrates that many or perhaps all of the degrees of freedom needed for a quantum phase gate, or other quantum information operations, are available. Existing work on classical and quantum information aspects of Laguerre-Guass beams, modes, and photons—no quantum dots involved—includes Refs. [54, 55, 56, 57, 58].

3.4.3 Conclusions

The numerical methods of Chap. 2 have allowed us to be the first (to our knowledge) to discover the classical spin-orbit coupling of paraxial modes. Through a number of numerical experiments using these methods, and through paraxial and electromagnetic theory, we have characterized the effect of this coupling, the mixing of different spin-OAM modes, in axisymmetric Fabry-Perot cavities with a planar Bragg mirror. The mixing is also expected to be found in other cavities possessing at least one Bragg mirror. As mentioned at the beginning of this chapter, we have found that the mixing occurs within pairs, and is governed by a single mixing parameter, α, the behavior of which is thoroughly investigated by numerical experiment. The mixing angle depends strongly on the effective birefringence of the Bragg mirror and is always zero for mirrors for which $r_s = r_p$. For Bragg mirrors, the mixing angle crosses from $45°$ to $-45°$ as $\epsilon_s - \epsilon_p$ is varied, producing a plot that can be fit by an arctangent or a sigmoid. Maximal mixing is the typical situation and the variation of α can be loosely interpreted as an anticrossing of an "S mode" and a "P mode". The qualitative behavior of the mixing angle is not affected by either the paraxiality parameter h or the resonance width $|2\text{Im}\,k|$, as long as both are sufficiently small; we may say that the mixing angle is zero order both in deviations from paraxiality and in deviations from perfectly reflective mirrors. The independence on h is in some sense surprising, as the birefringence experienced by the mode decreases to zero as the mode becomes more and more paraxial. However, the discussion of Sec. 3.1.3 shows that the "+" and "−" modes may mix to any degree in the paraxial limit, predicting the zero order dependence of α on h.

The observability of mixing depends on the observability of the degeneracy-breaking of the transverse mode families which are specified by (N, q). The observability for splitting requires cavities of sufficiently small length and high mirror reflectivity.

Fabry-Perot cavities which meet the observability requirements are likely to exist already, especially in VCSELs. Polarization measurements with such cavities should verify the characteristic mixing angle crossover from $-45°$ to $45°$, which we have measured with the two different numerical methods described in Chap. 2. Because dielectric mirrors are ubiquitous, it is to be expected that transverse degeneracy will routinely be lifted as smaller cavities become feasible; this will make it necessary to know how the vectorial eigenmodes are chosen. Furthermore, as optical resonator and laser devices for communication, computation, chemistry experimentation, nanoengineering, BEC, and other R&D applications get smaller and smaller, while the schemes of operation get more and more sophisticated, it seems inevitable that important uses of the controllable mixtures of Laguerre-Gauss modes will be discovered.

Chapter 4

SIMPLIFIED MODELS FOR SPIN-ORBIT COUPLING

In this chapter we discuss current progress in three different approaches to the calculation of the mixing angle using a basis consisting of two LGσ modes. As mentioned in Sec. 3.1.3, the argument for the linear mixing transformation (3.16) is expected to hold, in an appropriately loose sense, for the transverse mode patterns, not for the modes themselves. Attempting to apply the transformation in a rigorous sense to the modes themselves evokes the open-ended question: How should one choose the exact Gaussian mode parameters for the two basis modes? In particular, how should z_1 and ϕ_0 be chosen for these modes?[1] As will be seen in the next section, sensitivity to different problem formulations causes great difficulty in attempts to solve the cavity problem numerically, with a method similar to those of Chap. 2, using the two-mode basis. The difficulties in the direct numerical approach cast an ominous shadow on the *more* simplified (and more "physical") approaches discussed in Sec. 4.2 and at the end of the chapter. Nevertheless, these approaches, which begin from radically different starting positions, possess enough merit in their own right that they are included in this chapter.

Except where otherwise noted, the discussion here represents the individual, original work of the author.

---

[1]Here we are setting the origin $z = 0$ at the free-space waist locations of the mode envelopes (see the beginning of Sec. 3.1.4).

4.1 Direct Two-Mode Basis Solution

The most straightforward way to simplify the exact calculation and reproduce the mixing angle behavior is to directly solve the cavity problem using the mode basis consisting of the known modes $\mathrm{LG}_{p_+}^{|l_+|}\sigma^+$ and $\mathrm{LG}_{p_-}^{|l_-|}\sigma^-$ as given in Appendix A. The M1 boundary conditions are automatically satisfied in this method. The method is similar to the methods of Chapter 2. Some number of equations (nonlinear in this case) are constructed setting the appropriate fields ($E_\parallel$, E_ϕ, and $H_\perp$) equal to zero at some number of locations (rings) on the conducting mirror M2. As before, we can always set $\phi = 0$ for simplicity. The right hand side of each equation is zero, or should be. The complex squares of all of the equations are added together to form a target, Δ_t, which is minimized with respect to the unknowns, which are a subset of $\{\alpha, k, z_1, w_0\}$ (here k is taken to be real). The independent variables are those which specify the cavity, L, R, and k_d, as well as N, m and q, which identify which mixable mode pair should form the basis. A satisfactory solution should reproduce the mixing behavior, find α and k by minimizing, and find z_1 and w_0 by either minimizing or calculating using paraxial formulas such as those given in the next subsection. Such a solution has not been found, although, as will be shown shortly, the shape of the mixing curves have been reconstructed using the "cheat" that the value of k be set to the known k_C (or k_D) for each data point.

4.1.1 The Parameter z_1

We set the origin for the basis modes to be at the location of their waists, as discussed in the beginning of Sec. 3.1.4. With this origin, the surfaces of M1 and M2 are located at z_1 and z_d respectively, so that z_d and z_1 change if the waist location changes, while $z_d - z_1 \equiv L$ remains constant. In a numerical solution, z_1 can be set to a fixed value, such as 0, or used as a minimization parameter. There is also the option of calculating a value for z_1, in terms of k, using Gaussian formulas which are rigorously applicable in the paraxial limit to cavities with "simple" mirrors (mirrors of constant phase shift). To the author's knowledge, there is no conclusive argument for how z_1 should be treated for the basis modes, which have non-simple mirrors. Next we derive the calculation of z_1 for simple mirrors.

For a scalar Gaussian mode in a cavity with a M1 of constant phase shift ϕ_0, a relation between z_1, k and ϕ_0 can be established from paraxial theory. To estimate

z_1 in the paraxial limit, one can first obtain the Rayleigh range as $z_R = \sqrt{z_d(R - z_d)}$ from the last equation of (3.7). This expression can then be substituted into Eq. (3.12) with the replacements $z_w \to 0$, $z_2 \to z_d$, $z_1 \to z_1$, $\phi_1 \to \phi_0$, and $\phi_2 \to \pi$, resulting in

$$\arctan \frac{z_d}{\sqrt{z_d(R - z_d)}} - \arctan \frac{z_d - L}{\sqrt{z_d(R - z_d)}} = \frac{1}{N + 1}\left(kL - (q + 1)\pi + \frac{\pi}{2} + \frac{\phi_0}{2}\right).$$
(4.1)

Solving this transcendental equation for z_d and subtracting L yields z_1. When $z_1/L \ll 1$, a first order expansion yields

$$z_1 = -2\sqrt{L(R - L)}\left[\frac{kL - (q + 1)\pi + \frac{\pi}{2} + \frac{\phi_0}{2}}{N + 1} - \arctan\sqrt{\frac{L}{R - L}}\right].$$
(4.2)

Note that z_1 is extremely difficult to obtain numerically from our simulations, so that it is difficult to check the accuracy of Eqs. (4.1, 4.2). However, numerical checks on paraxial vector modes in cavities with r_s and r_p set to $e^{i\phi_0}$, for a handful of "general" ϕ_0 values, show that z_1, calculated by Eq. (4.1), is very close to zero (much less than a wavelength).[2] This result suggests two things: 1) the waist location is always at the M1 surface when M1 acts as a constant phase shift, and 2) the formula (4.1) is accurate for vector modes with the h values we have been considering ($h < 0.1$). (The first point is obviously true when M1 is a perfect conductor ($\phi_0 = \pi$), but it is not obvious to see how this arises for general ϕ_0.)

It is reasonable to attempt to use Eq. (4.1) to calculate z_1 for cavities with a Bragg mirror. The numerical eigenvalue k is used and ϕ_0 is set to be $\phi_0(k)$ for the Bragg mirror. The z_1 values for the C1, C2, and C3 data obtained from the approximate formula (4.2) are given in Appendix C. As can be seen, $|z_1|$ can become a significant fraction of λ.

In the results of the next subsection, we do include results for which z_1 is calculated this way. It seems that the other natural option, setting $z_1 = 0$, produces better numerical results. Minimizing with respect to z_1 has also been done, but has not been beneficial for the cases that we have tried, nor has it shed much light on the question of what z_1 "should" be, due to difficulties of the type discussed in the next subsection.

---

[2]When the non-paraxial corrections for pure LGσ modes are included from Ref. [46], the calculated value of $|z_1|$ drops further by one or two orders of magnitude.

4.1.2 Numerical Results

A program has been constructed allowing α, k, z_1, and w_0 to be fixed to a set value, to be scanned over a fixed interval, or to be among the minimization variables which the minimizing routine (from the GNU Scientific Library, GSL) alters as free parameters. In addition, z_1 can be calculated using a fast 1D root find with Eq. (4.1), and w_0 can be calculated from z_1 and k using the formula $w_0^2 = (2/k)\sqrt{z_d(R - z_d)}$ derivable from Gaussian mode theory (3.7). Any combination of these actions for any of the mode parameters $\{\alpha, k, z_1, w_0\}$ can be realized in the program, allowing a large number of options to be explored. The exact functions for r_s and r_p are used in order to avoid singularities in the numerical integrations over $(s_u + s_d)^{-1}$ and $(p_u + p_u)^{-1}$ which occur when $|r_{s/p}| = 1$. The mixing angle is calculated by attempting constructing the C mode $(\cos(\alpha)\mathrm{LG}_{p+}^{l+}\sigma^+ - \sin(\alpha)\mathrm{LG}_{p-}^{l-}\sigma^-)$ and minimizing Δ_t. Solutions where $45° < |\alpha| \leq 90°$ indicate that D mode has been found.

Additionally, any number of boundary equations can be set, each complex equation generally yielding two constraints. Of particular interest are boundary equations on the z axis ($\rho = 0$), since these boundary equations are the simplest and are not manifestly sensitive to deviations of the spherical M2 from the actual phase front shape. As the z axis is approached, the equations $E_\parallel = 0$, $E_\phi = 0$, and $H_\perp = 0$ approach interdependence, so that only a single boundary condition is obtained. However there exists a second independent z-axis boundary condition if curvature is taken into account. The two independent boundary conditions are

$$w_0 E_\rho(0, 0, z_d) = 0,$$

$$w_0 R \frac{\partial}{\partial \rho} H_z(0, 0, z_d) + w_0 H_\rho(0, 0, z_d) = 0, \tag{4.3}$$

(in either cylindrical or Cartesian coordinates). The factors of w_0 are included because the transverse fields are normalized so that $\int |E_T|^2 dA = constant$, meaning that $w_0 E_T$ does not monotonically vary with w_0. Numerically we find it makes little difference whether the factor of w_0 is included or not, even when minimizing with respect to w_0.

Figure 4.1 shows the two-basis version of the mixing curve for the C1 and C3 data. The results roughly agree with the correct mixing curves, although there is a significant departure from the $-45°$ asymptote. The four boundary conditions are the real and imaginary parts of the equations (4.3). At each point plotted, k_d was

Figure 4.1: Two-mode basis results for C1 (diamonds) and C3 (boxes) mixing curves compared with previous numerical results (C1 = solid, C3 = dashed).
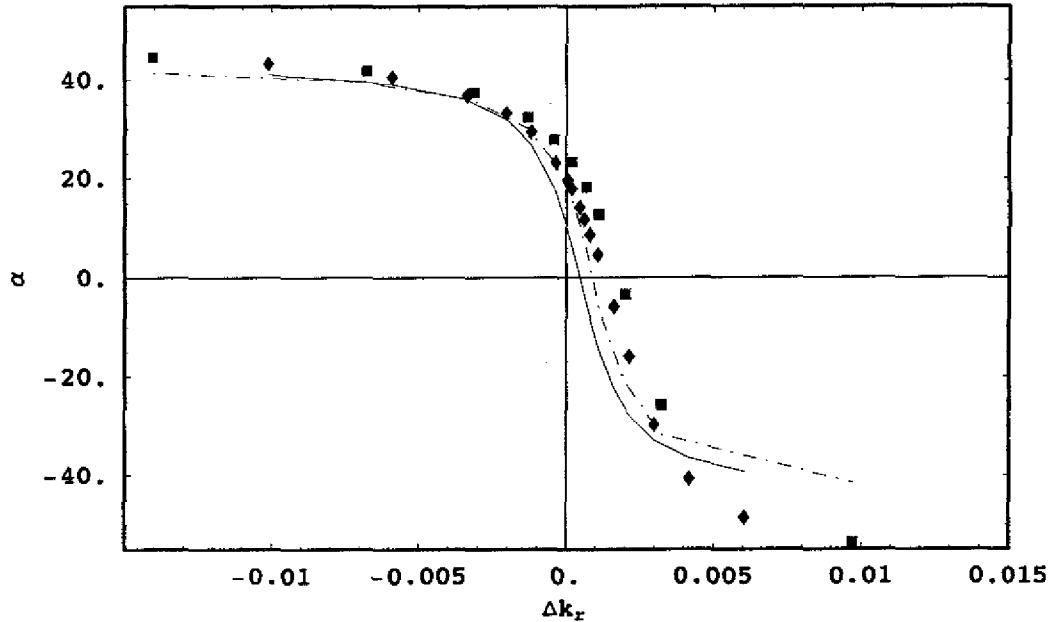
set to the design wavelength used in the previous C1 or C3 numerical experiments, z_1 was set to 0, k was set to the previously calculated value of k_C, and α and w_0 were minimization parameters.

As previously mentioned, setting k instead of finding it by minimization is a cheat, and result in Fig. 4.1 is not a true prediction. The difficulty with finding k by minimization is due to both a high sensitivity of the position of the global minimum to the parameters, and the presence of many artificial local minima. Perhaps a tailor-made minimization routine could overcome the latter problems, but the former problem is the most dominant and problematic. Changing the problem formulation to any of the other reasonable combinations of options for z_1 and w_0 moves not only the global minima, but in general moves the k value at which α is correct further away from k_C, so that the "cheating" mixing curve becomes worse and not better. Selecting or adding off-axis M2 boundary conditions seems to dramatically affect the global minima position as well.

Any significant improvement to the direct two-basis method itself will have to

involve some changes to the basis functions themselves. One possible improvement is to allow the two basis functions to have different values of z_i and w_0. Other improvements perhaps involve arguments for certain functional modifications of the basis functions themselves.

The next section gives a rather indirect two-basis method which does not involve boundary conditions at M2.

4.2 A Degenerate Perturbation Theory Approach

Because the cavity problem with a Bragg mirror problem involves both k-space boundary conditions and real space boundary conditions, we do not know of any existing framework in which to construct a perturbation theory for the electromagnetic field in the small quantity h, or even the quantity $\epsilon_s - \epsilon_p$, supposing it could be varied smoothly from zero. What we present here is an effective theory which can be loosely interpreted as the zero-order calculation of a degenerate perturbation theory in h using a two-mode basis. Our perturbation theory in $h = 2/(kw_0)$, successfully handles the difficulties that arise from having infinite wavevector or infinite waist radius (or both) for the unperturbed ($h = 0$) wavefunctions. While the argument behind the construction of the theory is obviously incomplete, the remarkable result indicates that *the methods used may be part of some unknown, complete theory*. Our theory predicts that, in the paraxial limit, the mixing angle is given by

$$\alpha \approx \frac{1}{2} \arctan[\zeta_{\text{th.}}(\epsilon_s - \epsilon_p)]. \tag{4.4}$$

The value of $\zeta_{\text{th.}}$ is arbitrary in that it depends on arbitrary functions; the value $\zeta \approx -18.5$ suggested by the numerical data of Fig. 3.11 might be obtainable as well as many other values. Setting the arbitrary functions to a value other than the simple constant 1 would be analogous to Copernicus' incorporation of small epicycles into his heliocentric theory to account for the fact that simple circular orbits of planets around the sun did not match observational data. However, there is reason to believe that the arbitrary functions would, in the posited complete theory, be specified in some way *by the parameters which are set by the location and radius of the curved mirror*. As it stands, our prediction of the *form* (4.4) is obtained *with little or no information* about M2. This is the principal asset of this theory: it isolates the effect of the Bragg mirror. The prediction of the form involves only ϵ_s and ϵ_p, characteristics

of the M1, and mild assumptions on the behavior of ϵ_s, ϵ_p, and kz_1, the last of which could conceivably depend on M2 via the discussion of Sec. 4.1.1.

4.2.1 The Zero-Order Prediction of a Degenerate Perturbation Theory

The basic procedure in degenerate perturbation theory follows below (see for example Sec. 5.2 of Ref. [59]).

1. Determine the degenerate unperturbed eigenkets

2. Construct the perturbation matrix V. The dimension of V is n if the degeneracy is n-fold.

3. Diagonalize the perturbation matrix. The base kets that diagonalize the V matrix are the correct zeroth-order kets to which the perturbed kets approach in the limit of the perturbation going to zero. The eigenvalues are first-order energy (or frequency) shifts.

For our situation, $g = 2$, λ is h, and the unperturbed eigenkets $|\psi_+\rangle$ and $|\psi_-\rangle$ will represent the $LG_{p_+}^{l_+}\sigma^+$ mode and the $LG_{p_-}^{l_-}\sigma^-$ mode with h taken to be 0, ignoring potential problems such as the fact that either k or w_0 or both must be infinite. (The infinite quantities ultimately cancel.)

In the next section we will construct the matrix V. To find the perturbed zero-order kets, which we can call $|C\rangle$ and $|D\rangle$, we can simply set $\langle C|V|D\rangle = 0$ to find α. The procedure works as follows:

$$\begin{aligned}
\langle C|V|D\rangle &= (\cos(\alpha)\langle\psi_+| - \sin(\alpha)\langle\psi_-|)V(\sin(\alpha)|\psi_+\rangle + \cos(\alpha)|\psi_-\rangle \\
&= \cos(\alpha)\sin(\alpha)(\langle\psi_+|V|\psi_+\rangle - \langle\psi_-|V|\psi_-\rangle) \\
&\quad + \cos^2(\alpha)\langle\psi_+|V|\psi_-\rangle - \sin^2(\alpha)\langle\psi_-|V|\psi_+\rangle .
\end{aligned} \tag{4.5}$$

We will find that $\langle\psi_-|V|\psi_+\rangle = \langle\psi_+|V|\psi_-\rangle$, so that setting $\langle C|V|D\rangle$ to zero yields

$$\tan(2\alpha) = \frac{2\langle\psi_+|V|\psi_-\rangle}{\langle\psi_-|V|\psi_-\rangle - \langle\psi_+|V|\psi_+\rangle}, \tag{4.6}$$

or

$$\alpha = \frac{1}{2} \arctan\left[\frac{2\langle\psi_+|\,V\,|\psi_-\rangle}{\langle\psi_-|\,V\,|\psi_-\rangle - \langle\psi_+|\,V\,|\psi_+\rangle}\right]. \tag{4.7}$$

Our theory will only give us the zero-order result, that is, the mixing angle α. It cannot be used to estimate the energy shifts, which are first order in V_0, the overall magnitude of V, which must be an arbitrary parameter in our theory.

4.2.2 The Perturbing Hamiltonian V and the Unperturbed Kets

By constructing a perturbation V that produces qualitatively reasonable results, we can validate the degenerate perturbation theory hypothesis. Our notation relies heavily on Sec. 3.1.4 and Appendices A and B.

Let us for the moment consider the entire Bessel wave basis, and not the reduced basis $\{\psi_+, \psi_-\}$. From 1D quantum mechanics we know that the wave function is small where the potential energy is large. If one wants $\psi(x) = \langle x \mid \psi\rangle$ to be small at a particular value of x, one makes $V(x) = \langle x|\,V\,|x\rangle$ very large. We want the quantities

$$e^{ikz_1 \cos\theta_k} S_k(\theta_k) - e^{-ikz_1 \cos\theta_k} r_s(\theta_k) S_k(\pi - \theta_k), \text{ and}$$

$$e^{ikz_1 \cos\theta_k} P_k(\theta_k) - e^{-ikz_1 \cos\theta_k} r_p(\theta_k) P_k(\pi - \theta_k) \tag{4.8}$$

to be small, since the M1 equations (2.3) set them to zero. We want these to be small for all θ_k, something which cannot be enforced once the basis is reduced to two modes. Nevertheless, we press on and construct a large "potential barrier", V, to force the "wavefunction" $\psi(x; \text{s/p}, \text{u/d}) = \langle x; \text{s/p}, \text{u/d} \mid \psi\rangle = (S/P)_{\text{u/d}}(\theta_k(x))$ to be such that the desired M1 equations are satisfied.[3] As usual, $x \equiv (\sqrt{2}/h)\sin\theta_k = (kw_0/\sqrt{2})\sin\theta_k$.

---

[3] Herein lies one of the fundamental difficulties of the construction of our theory: the idea of making expressions equal to 0 by incorporating them in a potential barrier V is unsound in our case because there is no unperturbed Hamiltonian H_0 containing a second derivative in x (or θ_k). If one looks at 1D QM problems, one finds that the reason the wavefunction is small at the edge of a barrier, and the reason that the wavefunction decays exponentially into a barrier, comes from the fact that kinetic part of the Hamiltonian has a second derivative in x. The exponential decay is a solution to a second order differential equation. We know of no justification in our problem for an H_0 which looks like $\mathrm{d}^2/\mathrm{d}x^2$. As it is, *our problem contains no known unperturbed Hamiltonian at all.* Perhaps the spatial M2 boundary conditions could be included as H_0 in an improved version of this effective theory.

Using the full Bessel basis, we construct V to be any positive-weighted projection onto the *compliment* of the subspace of all states whose wavefunctions obey the M1 equations. Using $r_{s/p} = e^{\phi_{s/p}(x)}$ we set $V = V_s + V_p$ where

$$V_p = V_0 \int_0^\infty dx\, x \gamma_p(x) \left[e^{-ikc(x)z_1} |x; p, u\rangle - e^{ikc(x)z_1} e^{-i\phi_p(x)} |x; p, d\rangle \right]$$

$$\otimes \left[e^{ikc(x)z_1} \langle x; p, u| - e^{-ikc(x)z_1} e^{i\phi_p(x)} \langle x; p, d| \right],$$

$$V_s = V_0 \int_0^\infty dx\, x \gamma_s(x) \left[(-i)e^{-ikc(x)z_1} |x; s, u\rangle - (-i)e^{ikc(x)z_1} e^{-i\phi_s(x)} |x; s, d\rangle \right]$$

$$\otimes \left[(i)e^{ikc(x)z_1} \langle x; s, u| - (i)e^{-ikc(x)z_1} e^{i\phi_s(x)} \langle x; s, d| \right]. \tag{4.9}$$

V_0 is some large positive number giving height to the barrier. Here γ_s and γ_p are allowed to be arbitrary positive functions of x, because any weighting at all of quantities that we want to be zero should be equally applicable. As the reader may immediately check, the hermitian V has been constructed so that, if the wavefunction $\langle x; s/p, u/d \mid \psi \rangle$ obeys the M1 equations, then the necessarily non-negative overlap, $\langle \psi | V | \psi \rangle$, drops to 0 (because V is a projection onto the orthogonal subspace). The converse of the if-then statement is also true. The small quantity h enters V through both

$$c(x) \approx 1 - (h^2/4)x^2 \tag{4.10}$$

and

$$\phi_{s/p}(x) \approx \phi_0 + (h^2/2)\epsilon_{s/p}x^2. \tag{4.11}$$

As previously noted, we will pay some special attention to the simple constant function $\gamma_s = \gamma_p = 1$.

The unperturbed states are obtained with the help of Eqs. (3.27, 3.25, 3.20) for

$h = 0$. The unperturbed wavefunctions are given by

$$\langle x; \mathrm{p}, \mathrm{d} \mid \psi_\pm \rangle = \pm \underbrace{\frac{(-i)^{m-1+2p_\pm}\zeta_\pm}{4\pi^{3/2}}\sqrt{\frac{p_\pm!}{(p_\pm + |l_\pm|)!}}}_{c_\pm}\,\underbrace{x^{|l_\pm|}L_{p_\pm}^{|l_\pm|}(x^2)e^{-x^2/2}\,e^{ikz_1}}_{f_\pm(x)},$$

$$\langle x; \mathrm{p}, \mathrm{u} \mid \psi_\pm \rangle = c_\pm f_\pm(x)e^{i\phi_0}e^{-ikz_1},$$

$$\langle x; \mathrm{s}, \mathrm{u/d} \mid \psi_\pm \rangle = \mp i\,\langle x; \mathrm{p}, \mathrm{u/d} \mid \psi_\pm \rangle. \tag{4.12}$$

Since V is the perturbation Hamiltonian, it should go to zero when h goes to zero. A glance at Eq. (4.9) perhaps does not make it clear that this is the case. However, one sees that $V(h = 0)$ is effectively zero because the action of V on either of the unperturbed reduced basis states (or any other states which obey the M1 equations at $h = 0$), yields 0. In other words $V(h = 0)|\psi\rangle = V(h = 0)(a|\psi_+\rangle + b|\psi_-\rangle) = 0$ for any a and b.

4.2.3   Calculation of the Inner Products

Here we calculate $\langle \psi_\pm | V | \psi_\pm \rangle$ and $\langle \psi_\pm | V | \psi_\mp \rangle$.

$$\begin{aligned}
\langle \psi_\pm | V | \psi_\pm \rangle = V_0 \int_0^\infty \mathrm{d}x\, x \\
\times \Big[\gamma_\mathrm{p}(x)\Big(e^{-ikc(x)z_1}\langle \psi_\pm \mid x; \mathrm{p}, \mathrm{u}\rangle - e^{ikc(x)z_1}e^{-i\phi_\mathrm{p}(x)}\langle \psi_\pm \mid x; \mathrm{p}, \mathrm{d}\rangle\Big) \\
\times \Big(e^{ikc(x)z_1}\langle x; \mathrm{p}, \mathrm{u} \mid \psi_\pm\rangle - e^{-ikc(x)z_1}e^{i\phi_\mathrm{p}(x)}\langle x; \mathrm{p}, \mathrm{d} \mid \psi_\pm\rangle\Big) \\
+ \gamma_\mathrm{s}(x)\Big((-i)e^{-ikc(x)z_1}\langle \psi_\pm \mid x; \mathrm{s}, \mathrm{u}\rangle - (-i)e^{ikc(x)z_1}e^{-i\phi_\mathrm{s}(x)}\langle \psi_\pm \mid x; \mathrm{s}, \mathrm{d}\rangle\Big) \\
\times \Big(ie^{ikc(x)z_1}\langle x; \mathrm{s}, \mathrm{u} \mid \psi_\pm\rangle - ie^{-ikc(x)z_1}e^{i\phi_\mathrm{s}(x)}\langle x; \mathrm{s}, \mathrm{d} \mid \psi_\pm\rangle\Big)\Big]. \tag{4.13}
\end{aligned}$$

Using Eqs. (4.10, 4.11, 4.12) yields

$$\begin{aligned}
\langle \psi_\pm | V | \psi_\pm \rangle = V_0 \int_0^\infty \mathrm{d}x\, x\Big[\gamma_\mathrm{p} c_\pm^* f_\pm e^{-i\phi_0}\Big(e^{ikz_1(h^2/4)x^2} - e^{-ikz_1(h^2/4)x^2}e^{-i(h^2/2)\epsilon_\mathrm{p}x^2}\Big) \\
\times c_\pm f_\pm e^{i\phi_0}\Big(e^{-ikz_1(h^2/4)x^2} - e^{ikz_1(h^2/4)x^2}e^{i(h^2/2)\epsilon_\mathrm{p}x^2}\Big)
\end{aligned}$$

$$+ \gamma_s c_{\pm}^* f_{\pm} e^{-i\phi_0} (\pm 1) \left(e^{ikz_1(h^2/4)x^2} - e^{-ikz_1(h^2/4)x^2} e^{-i(h^2/2)\epsilon_s x^2} \right)$$

$$\times c_{\pm} f_{\pm} e^{i\phi_0} (\pm 1) \left(e^{-ikz_1(h^2/4)x^2} - e^{ikz_1(h^2/4)x^2} e^{i(h^2/2)\epsilon_s x^2} \right) \Bigg]. \tag{4.14}$$

Each difference of exponentials in this expression is of order h^2. Since the result is proportional to the product of two of these differences, the final result has leading order h^4. Despite this, one can see that there is no need to expand $c(x)$ or $\phi_{s/p}(x)$ to order h^4. Thus we obtain

$$\langle \psi_{\pm} | V | \psi_{\pm} \rangle = \frac{h^4}{4} V_0 |c_{\pm}|^2 \int_0^{\infty} x^5 f_{\pm}^2(x) \left[\gamma_p(x)(\epsilon_p + kz_1)^2 + \gamma_s(x)(\epsilon_s + kz_1)^2 \right] dx. \tag{4.15}$$

We will now assume that, for some unknown reason, we must have $\gamma_s(x) = \gamma_p(x) = \gamma(x)$. This yields

$$\langle \psi_{\pm} | V | \psi_{\pm} \rangle = \frac{h^4}{4} V_0 \left(\epsilon_s^2 + \epsilon_p^2 + 2kz_1(\epsilon_s + \epsilon_p) + 2(kz_1)^2 \right) |c_{\pm}|^2 \int_0^{\infty} x^5 f_{\pm}^2(x) \gamma(x) \, dx. \tag{4.16}$$

A similar calculation reveals

$$\langle \psi_{\pm} | V | \psi_{\mp} \rangle = -\frac{h^4}{4} V_0 (\epsilon_s - \epsilon_p)[\epsilon_s + \epsilon_p + 2kz_1] c_{\pm}^* c_{\mp} \int_0^{\infty} x^5 f_+(x) f_-(x) \gamma(x) \, dx. \tag{4.17}$$

Checking with Eq. (4.12) reveals that $c_+^* c_- = c_+ c_-^*$ so that $\langle \psi_+ | V | \psi_- \rangle = \langle \psi_- | V | \psi_+ \rangle$ as asserted in Sec. 4.2.1.

4.2.4   Results of the Effective Theory

Inserting the results of the previous section into Eq. (4.6) yields

$$\tan(2\alpha) = (\epsilon_s - \epsilon_p) \left(\frac{(\epsilon_s + kz_1) + (\epsilon_p + kz_1)}{(\epsilon_s + kz_1)^2 + (\epsilon_p + kz_1)^2} \right) \left(\frac{2Q_0}{Q_+ - Q_-} \right). \tag{4.18}$$

Figure 4.2: "Prediction" of mixing angle for cavities C1 (stars), C2 (diamonds), and C3 (squares). The arbitrary function $\gamma(x)$ is set to unity. The correct (numerically calculated) mixing angles are shown in Fig. 3.11.
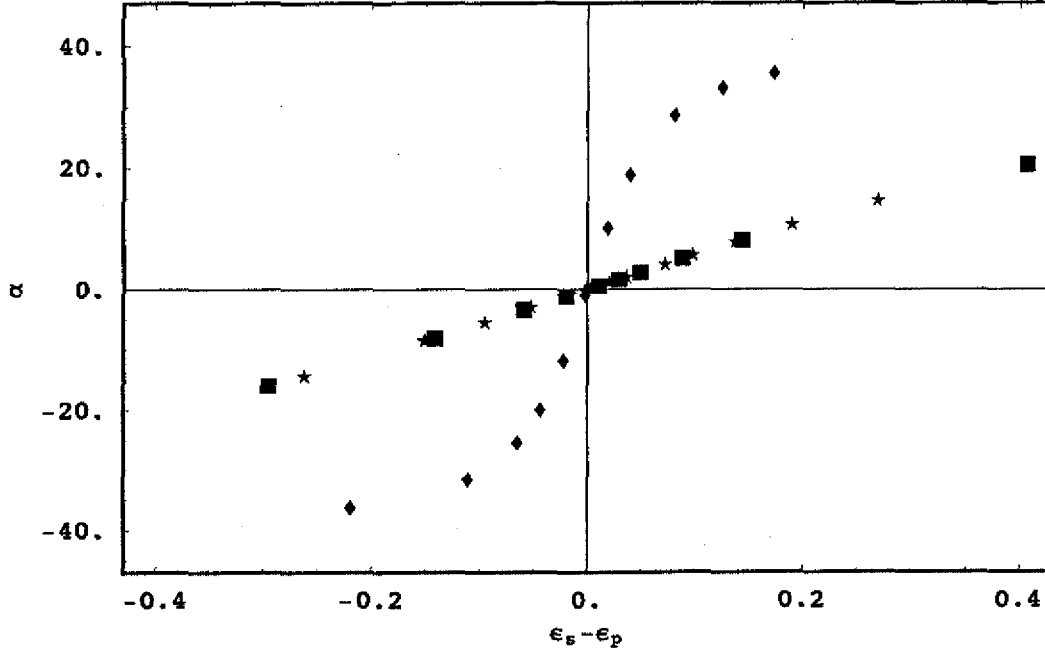
where

$$Q_0 = c_+^* c_- \int_0^\infty x^5 f_+(x) f_-(x) \gamma(x) \, dx,$$

$$Q_\pm = |c_\pm|^2 \int_0^\infty x^5 f_\pm^2(x) \gamma(x) \, dx. \tag{4.19}$$

Except for certain values of kz_1, (or more generally, certain functional dependencies of $kz_1(\Delta k_r)$), the middle factor of (4.18) varies relatively little (with Δk_r) compared to $\epsilon_s - \epsilon_p$, which sweeps through 0. Thus we may use the form given in Eq. (4.4) as an approximate form of α.

As one can see from (4.18), there is significant arbitrariness in $\zeta_{th.}$ from the arbitrary function $\gamma(x)$. The quantity kz_1 can be calculated from Eq. (4.2) with $k = (k_C + k_D)/2$ and q entered from the numerical results and $\phi_0(\Delta k_r)$ entered from stack calculations. The values of $\epsilon_{s/p}(\Delta k_r)$ also come from straightforward stack cal-

culations. (They are the same values used to obtain the abscissas in Fig. 3.11; in the present case however, the individual values ϵ_s and ϵ_p are used, not just their difference).

If we take $\gamma(x) = 1$, which is perhaps the most natural choice, the value of $\zeta_{th.}$, for each of the cavities C1, C2 and C3, is too small in magnitude and has the wrong sign. Figure 4.2 shows the "predicted" mixing curves, using Eqn. (4.18) with the middle factor included and z_i calculated via (4.1). The actual curve is given in Fig. 3.11. We note that setting z_i to zero instead of calculating it results in an increased slope of all three mixing curves, and the C2 curve becomes very close to the actual C2 curve (albeit with the incorrect sign).

At $\gamma(x) = x^{-4}$, the normalization relation (B.11) for $f_\pm(x)$ yields $Q_+ = Q_-$, while Q_0 remains finite. Thus $\zeta_{th.} \to \infty$ when $\gamma(x) \to x^{-4}$.

In conclusion, our effective theory produces the correct form of the solution but does not predict a specific quantitative function. It remains an open question whether some of the calculations in this section are found embedded in some more comprehensive paraxial mixing theory. The greatest value of the current theory is that it isolates the effect of a planar Bragg mirror and extracts the birefringence as the most important quantity.

4.3 Outlook for Simplified Models

The search for a simple explanation of the mixing behavior as a function of the effective birefringence of the Bragg mirror has been surprisingly difficult. Extensive searches of the literature have revealed no usable Hamiltonian or other framework for solving the *vectorial* cross-sectional mode patterns, and thus there is no known physically motivated framework for a perturbation theory. The zero-order perturbative behavior of the mixing angle indicates that a novel vectorial framework remains to be discovered.

As shown in Sec. 4.1, one formulation of the numerical solution with the two-mode basis reproduces the full numerical solution relatively well, but this formulation requires some numbers (the real k values) from the full solution to be put in as parameters. In order to improve this method, arguments must be developed that clarify the definitions of the unperturbed modes. A better understanding of how well the correct modes can be constructed using a two-mode basis would also be a helpful development.

The degenerate perturbation theory in h of Sec. 4.2 remarkably predicts the qualitative form of the dependence of α on $\epsilon_s - \epsilon_p$, but cannot give a quantitative prediction due to the existence of arbitrary functions in its formulation. This model essentially ignores the second cavity mirror M2. Perhaps there exists a very different formulation of the problem which would include the same theoretical kernel that our formulation creates, but without arbitrary functions.

Jens Nöckel has proposed another theory in which a generalized Born-Oppenheimer approximation (BOA) gives an alternate and interesting approach to solving scalar and vector paraxial cavity modes. The paraxial wave equation is not used, but the Gaussian mode family is derived using an effective radial potential which results from the BOA. Extending this theory to predict mixing behavior, however, currently appears to require the reduction of the basis modes to single Bessel waves. This effective angle approach is not effective because the Laguerre-Gauss modes are too broad and structured in θ_k to be well characterized by single effective polar angles.

One entirely different approach to the problem would be to find an extension of the perturbation methods of Erickson, Cullen, Yu and Luk [46] to mixed modes. At this time, however, it is completely unknown how to begin such an extension.

Another angle of attack, suggested by Nöckel, is to look for a way to incorporate the energy-k relation mentioned above into a perturbation-at-M1 approach, such as the method of Sec. 4.2. This approach, which might essentially be a perturbation theory in h, $(\epsilon_s - \epsilon_p)$, or both, would need to determine cavity energy perturbations due to changes in the momentum-space M1 equations.

While the search is still on for a simple explanation of spin-orbit coupling, the theoretical and numerical results presented in Chap. 3 give a thorough first investigation into the phenomenon. In particular, the discussions of Sec. 3.4 provide the general conclusions to our work on the classical spin-orbit coupling of paraxial modes. Perhaps the most important conclusion of the current chapter is that there seems to exist an electromagnetic perturbation phenomena, the simplest creating system of which is perhaps our model dome cavity, which does not succumb to any established electromagnetic perturbation theory framework. Thus this problem is of significant theoretical interest.

Chapter 5

OVAL DIELECTRIC RESONATORS, SEQUENTIAL TUNNELING, AND NONSPECULAR REFLECTION

This chapter first introduces whispering gallery (WG) modes in oval dielectric resonators, which is the primary topic of Chapters 5-7. The rest of the chapter discusses what happens at a single reflection of a ray representing an optical beam. (The larger goal is to represent circulating WG modes via the classical (or beyond classical) dynamics of the ray/beam.) The sequential tunneling model is a ray model in which loss of optical power occurs at each reflection, due to Fresnel transmission and/or effects of the curvature of the interface. We wish to determine whether we can add to sequential tunneling model nonspecular reflection corrections that could conceivably improve the usefulness of the ray model. The first and most important nonspecular correction that is discussed is the Goos-Hänchen effect. This effect shifts the starting position of the reflected ray along the surface. The second nonspecular phenomenon is due to what has been called Fresnel filtering in Ref. [60]. We will call this phenomenon the Fresnel kick, as in our application it changes the angle of the reflected beam so that $\theta_{refl} > \theta_{inc}$. The first part of the following chapter analyzes the feasibility of including the Goos-Hänchen effect into the sequential tunneling model. It appears that the feasibility question is answered in the negative, due to an approximation that is found to be generally invalid. The second part of Chapter 6 successfully employs

the GH effect in dome cavities. Chapter 7 begins with the exact wave equations, making more well-defined approximations to give a description of whispering gallery modes in oval dielectric resonators from a more concrete perspective. A numerical correspondence found in Chapter 7 supports the premise of sequential tunneling, that classical ray trajectories play a tangible role in WG modes.

5.1 Whispering Gallery Modes in Dielectric Oval Resonators

5.1.1 Introduction to Introductions to the Field

In the last three decades, the theoretical interest in bridging the gap between classical mechanical systems and their quantized counterparts has undergone exponential growth. The resulting body of work includes tighter connections to the old quantum mechanics, beautiful semiclassical theorems, such as the trace formulae of Gutzwiller [61] and Berry and Tabor [62], the field of quantum chaos, and an arsenal of problem-specific semiclassical methods and augmented classical models. This growth has been fueled in part by a surge of development in the field of dynamical chaos in the classical realm[1]. Trajectories in classical Hamiltonian systems can be either regular or chaotic, depending on the behavior of neighboring trajectories. Classical Hamiltonian systems are categorized as regular (or integrable) if all trajectories are regular, as chaotic if almost all trajectories are chaotic, or as mixed if regular and chaotic trajectories both have positive measure. (In mixed systems, the subsets of regular and chaotic trajectories form characteristic structures, called tangles, of amazing mathematical richness, in phase space.) For quantum systems, the fact that the Schrödinger equation is linear means that quantum evolution cannot be chaotic: there is no extreme sensitivity to the initial state of a wavepacket. The questions arises: When a chaotic or mixed classical system is quantized, what characteristics are imparted to the solutions of the quantum system? The majority of the many developments in this area

---

[1]The development of the computer has facilitated the rebirth of chaos in the 1970s and 1980s. Hamiltonian chaos, however is quite old, although the word "chaos" was not used. Poincaré published the construction of the chaotic homoclinic (or heteroclinic) tangle in 1899 [63]. In the 1920s Birkhoff constructed the Birkhoff normal form in a partially successful attempt to "transform away" chaos-generating nonlinear resonances. The important KAM theorem of Kolmogorov, Arnold, and Moser was developed in the 1940s, 1950s, and 1960s, and this work itself was built on previous knowledge of chaos in twist maps.

involve the distribution of energy eigenvalues: chaos leads to level repulsion, causing "spectral rigidity". Many of the later developments are focusing on the eigenfunctions themselves (e.g. [64, 65]). Included in these is a collection of studies of optical modes (in place of quantum eigenstates) in small resonators. In particular, oval dielectric resonators, which classically correspond to oval billiards, can have a very practical combination of properties, namely high Q and directional emission. The directional emission is manifestly due to the mixed phase space of the classical billiards [66, 67].

We do not (and cannot) give full introductions to classical Hamiltonian dynamics, semiclassical theory, or quantum chaos. An excellent graduate level textbook on classical mechanics, which leads the reader up to modern chaos studies, is José and Saletan [68]. (The popular book by Goldstein et al. [69] does poorly in this area, while Landau and Lifshitz [70] shamefully omits then entire existence of nonlinear dynamics, which was fathered by Poincaré.) An unsurpassed coverage of the mathematics and geometry of classical mechanics is given in Arnold [71]. Two excellent books describing Hamiltonian nonlinear dynamics are Lichtenberg and Lieberman [72] and Tabor [73].

It is more difficult to find good introductions to semiclassics or quantum chaos. Several books attempt to give the entire classical-semiclassical-quantum transition; one of the better of these advanced texts, in the author's opinion, is Ozorio de Almeida [3]. Quantum chaos spectra, their estimation by semiclassical trace formula, and their modeling by random matrix theory, are best introduced in other books, such as [4, 74]; we will not be interested in spectra or in results that only apply to fully chaotic systems. Our interests here lie in open, optical, mixed systems—the least amenable to most of the known classical$\leftrightarrow$semiclassical$\leftrightarrow$(quantum/wave) apparatuses. For introductions most suited to our applications, the reader is referred to the introductory material in the Ph.D. dissertations [75, 35, 34], all of which are easily obtained on the Web at the time of this writing. In this second part of the dissertation, much of the background theory will necessarily be covered by reference, rather than explanation.

5.1.2 Resonator and Billiard Model

As mentioned before, the simple classical picture of an oval resonator is the oval billiard: a point particle undergoing specular reflection[2] from the inside of an oval surface. In our oval resonator, we use a scalar wave in 2D to represent the optical field. This simplification is exact in the situation of an ideal cylindrical dielectric

---

[2]The term "specular reflection" simply means standard geometric reflection with $\theta_{refl} = \theta_{inc}$.

resonator, as shown in Fig. 1.1. The simplification is also exact (see section 8.7 of [8]) for non-infinite cylinders such as the cavities of microdisk lasers. For 3D oval resonators, the scalar wave simplification is an approximation that appears to be valid for some equatorially confined modes .

The wave equations and expansions, reprinted from Eqs. (1.2, 1.3, 1.4), are

$$\nabla^2 \psi_{\mathrm{i}} + n^2 k^2 \psi_{\mathrm{i}} = 0,$$
$$\nabla^2 \psi_{\mathrm{e}} + k^2 \psi_{\mathrm{e}} = 0, \tag{5.1}$$

$$\psi_{\mathrm{i}} = \sum_{m=-\infty}^{\infty} a_m J_m(knr)e^{im\phi}, \tag{5.2}$$

$$\psi_{\mathrm{e}} = \sum_m b_m H_m^{(1)}(kr)e^{im\phi}. \tag{5.3}$$

Note that there are no incoming waves in the resonance problem and we have switched from using ρ to r. From Eq. (1.5) the boundary conditions are

$$\sum_m a_m J_m(knr_G(\phi))e^{im\phi} = \sum_m b_m H_m^{(1)}(kr_G(\phi))e^{im\phi},$$
$$n \sum_m a_m J_m'(knr_G(\phi)) = \sum_m b_m H_m^{(1)\prime}(kr_G(\phi))e^{im\phi}. \tag{5.4}$$

For a circular boundary, $r_G(\phi)$ is constant, which results in modes with a single value of m, which is associated with angular momentum.

This is the "quantum", or wave, version of the problem. The simplest classical, or ray, version of the problem, is specular reflection in the billiard of the same shape. This version can be augmented to include, at least, an attenuating amplitude for the trajectory (sequential tunneling models), and non-specular reflection (of which the Goos-Hänchen effect is the most important type). Both of these augmentations are motivated by idea that the ray represents an optical beam, and both are related to the fact that the resonator we wish to model is bounded by a dielectric interface, not a conducting surface.

The unaltered billiard problem is characterized by an area preserving map (a canonical transformation) from one bounce point in phase space $(s_i, \sin \chi_i)$ to the next $(s_{i+1}, \sin \chi_{i+1})$. The position variable, s, denotes the boundary position (distance along the billiard boundary), modulo the billiard circumference; $\sin \chi$, where χ is the incident (and reflected) angle at the bounce point, is the momentum canoni-

cally conjugate to s. One can visualize a trajectory in phase space by plotting each bounce point i on a plot with abscissa s and ordinate $\sin \chi$. This plot turns out to be a Poincaré surface of section (see Sec. 1.1) of the full Hamiltonian billiard problem, which has four degrees of freedom (two momentum variables and two position variables) An explanation of these facts can be found in Ref. [75].

Since the equation of the billiard curve is usually parameterized by the angle ϕ instead of the curve length s, the surface of section (SOS) used in practice has abscissa ϕ. This plot is technically not a SOS, and the map from point to point is not area-preserving. Nevertheless, since ϕ increases monotonically with s, the appearance of the SOS is altered little and is just as useful for visualization and analysis of the Hamiltonian system.

Figure 5.1 shows trajectories in real space and phase space for a noncircular (quadrupolar) billiard. The circular billiard is integrable, or regular, while the quadrupolar billiard is nonintegrable and mixed, meaning that it contains both regular and chaotic trajectories. Some other nonintegrable billiards, such as the stadium, are fully chaotic, but "in general" billiards that one might draw with a pencil are mixed. The ellipse, featured in later chapters, is the most general noncircular billiard which is integrable. Each of the rare, periodic trajectories shows up as a set of discrete points on the SOS. Each of the quasiperiodic (regular) trajectories lie on (are dense on) a set of 1D curves on the SOS. For mixed systems, these curves are called KAM curves, and exist either as island chains, or as unbroken KAM-Lazutkin curves in the whispering gallery region near the top of the SOS. Finally, each of the chaotic trajectories fill up (are dense on) a multiply connected 2D region of positive measure in the SOS plot. Explanations of this topology and how it is created upon perturbation of an integrable system can be found in [72, 75].

In Refs. [1, 34], Tureci et al. established a successful and accurate Gaussian-beam-based theory of bouncing ball modes in dielectric optical resonators. The approach that Tureci used is based on linearized ray-optic reflection behavior which is not applicable to the high-angle reflections encountered in whispering gallery modes. For WG modes, Nöckel [75] has developed a semiclassical quantization using a particular curve in phase space, the approximate adiabatic invariant curve, as the domain of classical trajectory. The semiclassical quantization predicts values of $\operatorname{Re} k$ but does not directly predict resonance widths. To incorporate resonance widths, Nöckel used a classical sequential tunneling model in conjunction with the adiabatic curve. The resulting predictions for resonance width are in some cases surprisingly inaccurate (off
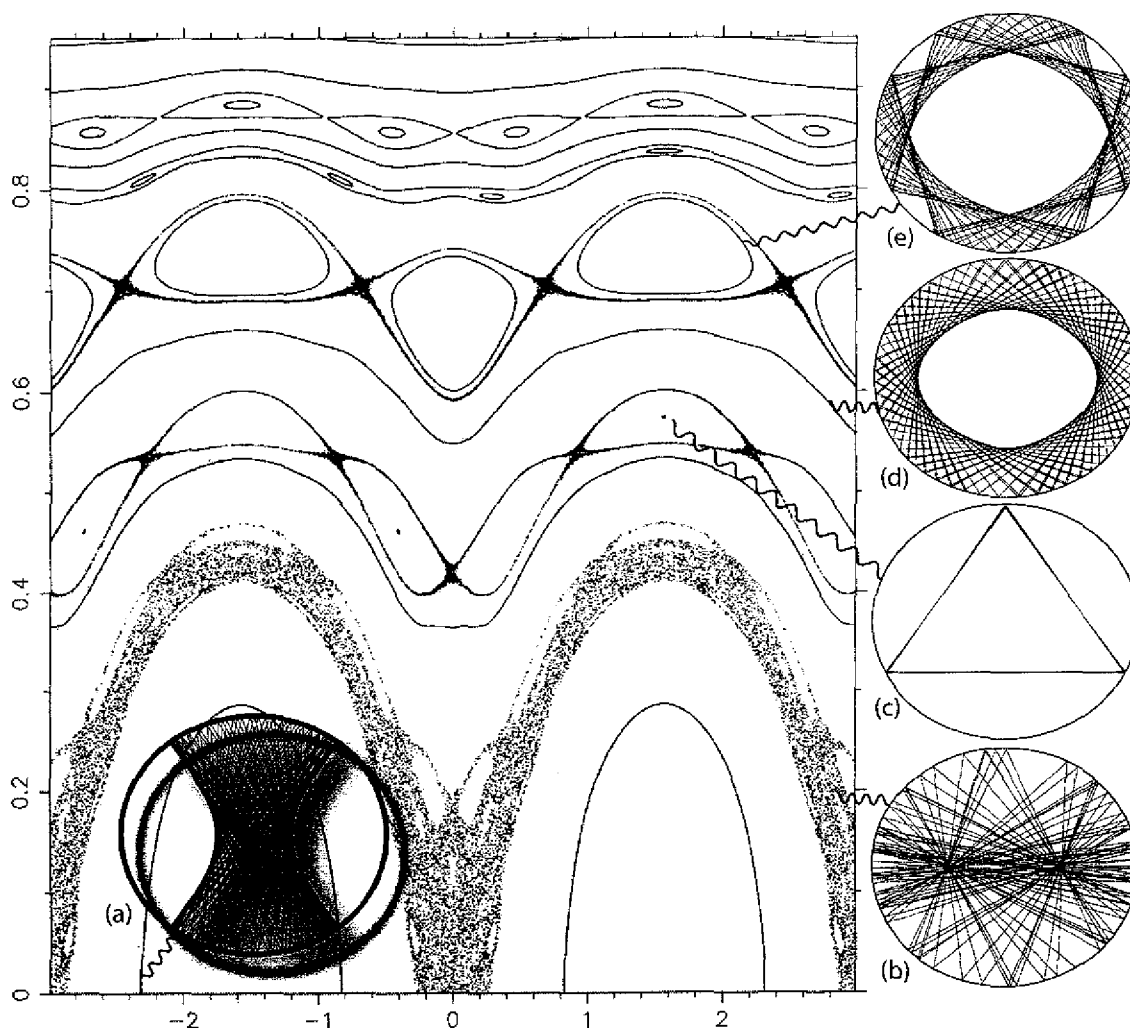
Figure 5.1: Surface of Section (SOS) for a quadrupole billiard of shape $r = 1 + \epsilon \cos(2\phi)$ with $\epsilon = 0.0526316$. The x axis is ϕ and the y axis is $\sin\chi$. Sixteen different trajectories are plotted, each one having run 20000 bounces. The locations of the two, three, four, five, and six-bounce islands can be seen. The three fuzzy phase space trajectories are chaotic separatrix orbits. Five of the trajectories are connected to real space pictures of the trajectories containing 100 bounces each. (a) A quasiperiodic regular orbit in the two-bounce stable island. (b) A chaotic trajectory surrounding the two-bounce island. (c) A quasiperiodic regular trajectory that is very close to one of the two primary stable three-bounce periodic orbits. (d) A KAM or KAM-Lazutkin curve (quasiperiodic and regular). Note the well-defined caustic. This type of orbit is associated with whispering gallery modes. (e) A quasiperiodic regular orbit in the four-bounce stable island.

by an order of magnitude or more) for moderately deformed oval resonators. In this chapter and the next we seek to determine whether the incorporation of nonspecular reflection, which is associated with the finite width of the bouncing beam, could cause a significant improvement in the sequential tunneling scenario.

5.2   The Sequential Tunneling Model

The sequential tunneling model assigns a power reflectance $R \leq 1$ to each bounce of a classical trajectory inside an oval resonator, with the intention of estimating the resonance widths $\gamma = -2\text{Im}\,k$ of the quasimodes. The easiest way to assign R of course is to treat the ray as a plane wave and approximate the curved dielectric boundary as being locally planar. The solution in this case is $R = |r_s|^2$ with the field reflection coefficient r_s being given by the Fresnel formula for s polarization,, shown here with $n_{\text{incident}} = n$, $n_{\text{transmitted}} = 1$:

$$r_s = \frac{n \cos \chi - \cos \chi_t}{n \cos \chi + \cos \chi_t}. \tag{5.5}$$

Since $\cos \chi_t = (1 - n^2 \sin^2 \chi)^{1/2}$, r_s is real for $\chi \leq \chi_c = \arcsin(1/n)$ and is on the unit circle for $\chi \geq \chi_c$. Thus, when χ is greater than the critical angle χ_c, there is total internal reflection (TIR). When the trajectories, which (one claims) support a given mode, have $\chi > \chi_c$ at every bounce, the planar reflection model just described predicts $\gamma = 0$, in contradiction with exact calculation. Most practical applications involve WG modes that are classically bounded in regions of high $\sin \chi$ due to the Lazutkin's theorem [75]. A useful sequential tunneling model must go beyond Fresnel's formulas and yield a decent approximation for the resonance widths of the modes of interest. To put it obliquely, the sequential tunneling model should include tunneling. The interpretation of a TIR ray "tunneling" out of an oval resonator is discussed in chapter 8 of [75] and in [76]. The essence of the idea is that a circular cavity problem can be reduced to a Hamiltonian-like problem for a radial "wavefunction" in an effective potential which has a minimum at the radius of the circle, R_c, and has a low energy region at large r which is accessible via tunneling.

An accurate sequential tunneling model was created by Nöckel [75] and later extended by Hentschel and Schomerus [77, 78, 79]. To assign a power reflectance R (and a power transmittance T) to a given ray reflection at a point with local radius of curvature R_l, the problem of a wave in a *circular* resonator of radius $R_c = R_l$ is

considered. The ideal situation is to find a mode for the posited circular resonant cavity that has the same real wavenumber $\operatorname{Re} k$ and the same angular momentum as the ray. The resonant width of this mode can then related to R by answering the question: If the ray were continually bouncing in the circle with radius R_1, what R would give the trajectory an average power decay of

$$P(t) = e^{-2ct|\operatorname{Im} k|} ?$$ (5.6)

In general there is no resonance state in this particular circular resonator that has either the correct $\operatorname{Re} k$ or the correct angular momentum, as both of these values are discrete for the wave problem. However, there is a natural way to relax restrictions and allow these parameters to be set at will.

Before relaxing any restrictions, the resonance problem for a circular resonator quantizes m to some integer. Complex k is quantized by the complex resonance equation (see (1.9))

$$n J_m'(nkR_c) H_m^{(1)}(kR_c) = J_m(nkR_c) H_m^{(1)\prime}(kR_c),$$ (5.7)

which simplifies, by recursion relations, to

$$J_m(nkR_c) H_{m-1}^{(1)}(kR_c) = n J_{m-1}(nkR_c) H_m^{(1)}(kR_c).$$ (5.8)

Consider a particular mode with its eigenvalue k and angular momentum quantum number m. (See Appendix F, where it is shown (with one physical assumption) that the physical angular momentum of a single z-polarized light quanta for a dielectric (or conducting) infinite circular cylinder problem is $m\hbar$.) It is straightforward to assign a trajectory in the circle billiard to this mode. In the ray picture the ray represents a single particle with momentum $\hbar n \operatorname{Re} k$. The angular momentum is calculated with respect to the origin at the center of the circle, and is

$$L_{\mathrm{ray}} = \hbar R_c n \operatorname{Re} k \sin \chi.$$ (5.9)

Equating $\hbar m$ with L_{ray}, we see that the mode is associated with the quasiperiodic (or

perhaps periodic) orbit having[3]

$$\sin \chi = m/(R_c n \mathrm{Re}\, k).$$ (5.10)

Imagine that we are considering a trajectory (in the circle) that exactly corresponds to an actual mode, via (5.10). From geometry, the length of a single ray (segment) of this trajectory is $2R_c \cos \chi$, which takes time $t_{\mathrm{ray}} = 2(n/c)R_c \cos \chi$ to traverse. In the sequential tunneling model, the ray power decreases by a factor of R every time step t_{ray}, yielding the smoothed dependence

$$\begin{aligned}
P_{\mathrm{ray}}(t) &= R^{t/t_{\mathrm{ray}}} \\
&= \exp\left[\frac{\ln(R)}{2nR_c \cos \chi} ct\right].
\end{aligned}$$ (5.11)

Comparison with (5.6) yields

$$R = e^{-4|\mathrm{Im}\, k| n R_c \cos \chi}.$$ (5.12)

For a bounce point in a noncircular billiard, R_c is replaced with R_l.

We now return to the problem of solving for $\mathrm{Im}\, k$. Since we wish to apply the sequential tunneling model to noncircular billiards, each bounce will have a different $\sin \chi$ and R_l, but the same wavenumber, $\mathrm{Re}\, k = \omega/c$. This means that the angular momentum quantity $n\mathrm{Re}(k)R_l \sin \chi$ is not generally an integer. The first restriction we relax, then, is that m be an integer. We assign

$$m = n\mathrm{Re}(k)R_l \sin \chi,$$ (5.13)

and use this m in Eq. (5.7). At this point we are essentially dropping the periodic boundary condition on ϕ in the circular billiard while retaining the boundary condition on r. The remaining boundary condition (5.7) is a complex equation that still sets both $\mathrm{Re}\, k$ and $\mathrm{Im}\, k$, while we wish to set $\mathrm{Re}\, k$ as a input parameter. The best that can be done is to use the fact that the solutions of $\mathrm{Re}\, k$ are spaced closely enough together that one can, in a rough approximation, ignore the quantization and simply plug the desired value of $\mathrm{Re}\, k$ into an expression (to be derived) that yields $\mathrm{Im}\, k$ as a function of $\mathrm{Re}\, k$ and the real number m.

---

[3]The entire orbit/trajectory can be associated with a single χ value for a circular billiard of course, since χ is a constant of the motion.

We assume we are interested in the case

$$|\mathrm{Im}\,kR_1| \ll 1. \tag{5.14}$$

In section 8.4 of Ref. [75], the formula

$$\mathrm{Im}\,kR_1 = \frac{2}{\pi(1-n^2)\mathrm{Re}\,kR_1}\frac{1}{J_m^2(\mathrm{Re}\,kR_1)+Y_m^2(\mathrm{Re}\,kR_1)} \tag{5.15}$$

is developed[4] by a leading order expansion in $\mathrm{Im}\,kR_1$ when m and $\mathrm{Re}\,k$ are given. Inserting this equation into (5.12) yields

$$R = \exp\left[-\frac{8}{\pi}\frac{n}{n^2-1}\frac{1}{\mathrm{Re}\,kR_1}\frac{\cos\chi}{J_m^2(\mathrm{Re}\,kR_1)+Y_m^2(\mathrm{Re}\,kR_1)}\right]. \tag{5.16}$$

The sequential tunneling model is thus completed for s polarized light. The classical ray is traced from bounce to bounce and its "power" is multiplied by $R(\sin\chi, R_1(s); \mathrm{Re}\,k)$ at every bounce. Once the power of the ray reaches a certain threshold, say $0.01P_{\mathrm{ray}}(0)$, the trajectory can be terminated.

The sequential tunneling model can be used when $\sin\chi$ is both above and below $\sin\chi_c = 1/n$, although the assumption $|\mathrm{Im}\,kR_1| \ll 1$ may fail below the critical angle, where direct (Fresnel) transmission takes place. The curve shown in Figure 5.2 shows $\ln[-\mathrm{Im}\,kR_1]$ using (5.15) with $\mathrm{Re}\,k = 8.0$, $R_1 = 10$, $n = 1.51$. The crosses are the exact mode solutions for the circle resonator (with integral m and $R_c = 10$) that have $\mathrm{Re}\,k$ within 1% of 8.0 (e.i. $7.92 < \mathrm{Re}\,k < 8.08$). The $\sin\chi$ value for the exact mode solutions comes from $\sin\chi = m/(n\mathrm{Re}\,kR_c)$. The agreement between exact solutions and the approximation is seen to be quite good. These modes solutions come in groups which are numbered on the plot by how many radial nodes exist in the resonance state. This number has the same range and meaning as p for the Laguerre-Gauss mode patterns. An exact mode with two radial nodes is shown in Fig. 5.3.

The *approximate* relative error incurred strictly from ignoring the quantization of $\mathrm{Re}\,k$ is shown in the inset of Fig. 5.2. Estimating the length of a trajectory's round trip as $2\pi R_1$, one can see that the spacing of $\mathrm{Re}\,kR_1$ solutions is approximately

---

[4]The original formulation has the quantity $[J_m(x)Y_{m-1}(x) - J_{m-1}(x)Y_m(x)]$ in the numerator, with $x \equiv \mathrm{Re}\,kR_1$. This quantity turns out to be $2/(\pi x)$ [77]. Also, for complicated reasons, the factor $1/(1-n^2)$ is sometimes approximated as $(1/2n)\ln[(n-1)/(n+1)]$ by Nöckel and Hentschel; we have found it is more accurate to use $1/(1-n^2)$.

Figure 5.2: Plot of $\ln[-\mathrm{Im}\,kR_{1/c}]$ vs. $\sin\chi$. The solid line uses Eq. (5.15) and the crosses indicate true mode solutions. The inset gives a continuous estimate of the relative error of $\mathrm{Im}\,kR_1$ (this is *not* a semilog plot). See text for description.
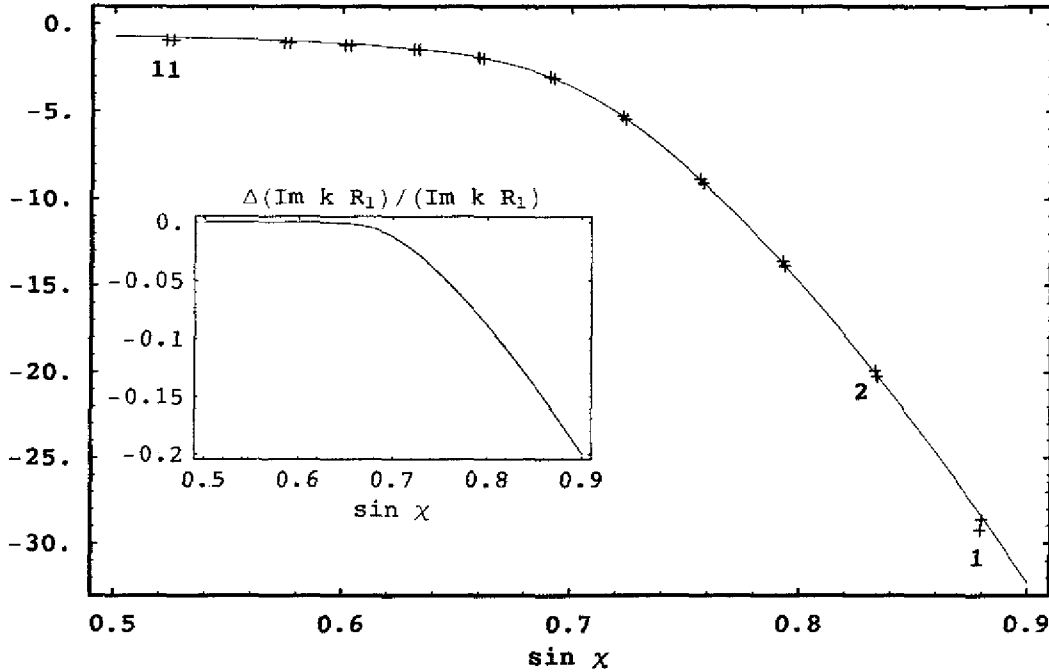
$\Delta(\mathrm{Re}\,kR_1) = 1$. The approximate error in $\mathrm{Im}\,kR_1$ from ignoring the quantization is therefore

$$\Delta(\mathrm{Im}\,kR_1) \approx \frac{\mathrm{d}\,\mathrm{Im}\,kR_1}{\mathrm{d}(\mathrm{Re}\,kR_1)} \times \frac{\Delta\mathrm{Re}\,kR_l}{2}$$

$$\approx \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{2}{\pi(1-n^2)x}\frac{1}{J^2_{nx\sin\chi}(x)+Y^2_{nx\sin\chi}(x)}\right]_{x=\mathrm{Re}\,kR_1}. \tag{5.17}$$

As can be seen in the plot, the relative error $|\Delta(\mathrm{Im}\,kR_1)|/(\mathrm{Im}\,kR_1)$ is below 20% for $\sin\chi < 0.9$.

An entirely different approach to sequential tunneling was developed by Hentschel [78, 79]. This model treats the ray as a Gaussian beam which intersects the boundary over a finite arc and also includes the Goos-Hänchen (GH) effect to displace the point of reflection along the arc, thus changing the incident angle. In this model, the GH effect, combined with the curvature of the billiard, shifts the range of incident angles. This modified range is integrated against the Fresnel transmission of each angle,

Figure 5.3: Close-up of the circle resonator mode for $n = 1.51$, $R_c = 10$. The mode has two radial nodes and corresponds to the lower of the two crosses labeled "2" in Fig. 5.2 and has $m = 101$, $k = 8.021344 - i1.63 \times 10^{-10}$.
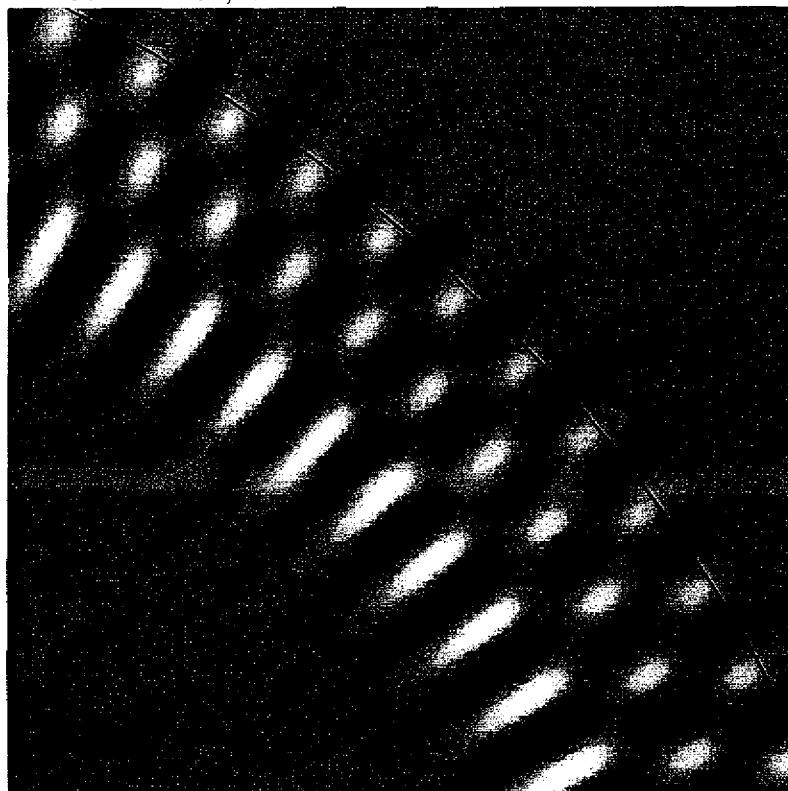
producing a modified average transmission coefficient for the beam/ray. Although this model is not as accurate as the one described above (which does not use the Fresnel formulas), the fact that the GH effect is used as a partial explanation for tunneling in [79] prompted hopes that the GH effect might be an important factor in dielectric resonators. The hope is that discrepancies in the resonance widths, found between exact solutions and the sequential tunneling model solutions (using either Nöckel's or Hentschel's model for transmission), might be explained by a further incorporation (not used in [79]) of the GH effect: directly including the GH shift in the ray dynamics.

We next discuss and calculate the Goos-Hänchen effect itself.

5.3 The Goos-Hänchen Effect: The Artmann Formulation

The GH effect is a parallel shift (in the plane of incidence) of a reflected beam of light (or other wave) away from its geometric (ray) position. The GH effect, is one of several known *nonspecular* corrections to beam reflection (or transmission). The other nonspecular effect that we will discuss is the Fresnel kick, the change in angle of the reflected beam, due to preferential transmission of the plane wave components (of the incident beam) having lower angles of incidence. A unifying derivation of these two nonspecular effects as well as two others are found in an article by Tamir [80]. The derivation only holds, however, under certain conditions which may not be met in our applications. These four effects discussed in Tamir are all changes (upon reflection) that occur to the Gaussian beam parameters describing the beam *in the plane of incidence*. There are also four (even) smaller effects on the Gaussian beam parameters describing the beam (shift, direction, semimajor/minor width, and waist location along the beam) in the direction perpendicular to the plane of incidence, which are summarized by Nasalski [81].

The nonspecular reflection effects are related to nonstandard transmission effects. The Fresnel kick is accompanied by a much larger deviation in angle (from that predicted by Snell's Law) in the transmitted beam [60]. Another nonspecular transmission effect, which is loosely conjugate to the transverse shift effect discussed by Nasalski, is the optical Magnus effect, a complicated parallel shift, perpendicular to the plane of incidence, of a transmitted, circularly polarized beam crossing a dielectric interface. (This sideways, polarization-dependent shift, predicted and discovered in the early 1990s by Zeldovich and Liberman (see [37] and references therein), is yet another potential candidate for explanation of the paraxial spin-orbit coupling of Chapters 3 and 4.)

The standard analytical explanation for the GH effect was given in by Artmann [82] soon after the experimental discovery in 1947 by Goos and Hänchen [83]. The simple derivation runs as follows.[5] A beam in medium A of refractive index n_A encounters a planar dielectric interface with angle of incidence χ and undergoes total

---

[5]Since the simple unification of four nonspecular effects by Tamir [80], the four effects should really be presented together. However, Tamir's presentation (in two pages) is almost impossible to improve upon, so we simply advise the reader to obtain this reference after reading the current section. (One improvement: in Eq. 18 the quantity $\text{Re}[(x_r - L)/w_f]$ should be $\text{Re}[(x_r - L)^2/w_f^2]$.)

Figure 5.4: The local coordinate systems of the central rays k_{inc} and k_{refl}.
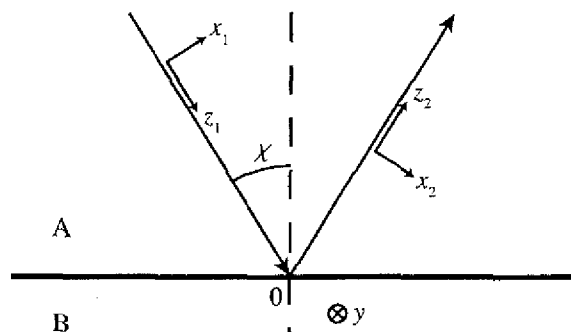
internal reflection. (To have TIR one must have $n_A > n_B$ and $\sin \chi > n_B/n_A$.) We set up two sets of coordinates (x_1, z_1) and (x_2, z_2) as shown in Fig. 5.4. The origin of both sets is at the intersection of the center of the incident beam with the interface. The oriented z_1 axis points along k_{inc} and the oriented z_2 axis points along the geometrically predicted k_{refl}. The x axes are each perpendicular to their respective z axes, with the orientation of x_2 chosen such that x_1 and x_2 coincide for $\chi = 0$.

The incident beam is decomposed into plane waves, each specified by its direction (k_{x1}, k_{z1}) and its weight $\tilde{\psi}(k_{x1})$, with the conditions $k_{x1}^2 + k_{z1}^2 = k_A^2 \equiv n_A^2 k^2$ and $k_{x1} \ll k_{z1}$. The reflected beam is likewise decomposed into plane waves of direction (k_{x2}, k_{z2}). One can see that the reflected plane wave of direction (k_{x2}, k_{z2}) gets weight $r_{\text{s/p}}[\theta_k = \chi + \arcsin(k_{x2}/k_A)]\tilde{\psi}(k_{x2})$, where $r_{\text{s/p}}$ is the plane wave reflection coefficient (this works for a dielectric stack as well as a single dielectric interface). The spatial dependence of the incident and reflected beams are thus

$$\psi_{\text{s/p inc}}(x_1, z_1) = \int_{-k_A}^{k_A} f_{\text{s/p}}(k_x)\tilde{\psi}(k_x) \exp[i(\sqrt{k_A^2 - k_x^2}z_1 + k_x x_1)]\, dk_x, \tag{5.18}$$

$$\psi_{\text{s/p refl}}(x_2, z_2) = \int_{-k_A}^{k_A} f_{\text{s/p}}(k_x)r_{\text{s/p}}(\theta_k(k_x))\tilde{\psi}(k_x) \exp[i(\sqrt{k_A^2 - k_x^2}z_2 + k_x x_2)]\, dk_x, \tag{5.19}$$

where

$$f_{\text{s}}(k_x) = 1,$$
$$f_{\text{p}}(k_x) = \sqrt{1 - (k_x/k_A)^2},$$

$$\psi_s(x_i, z_i) \equiv \boldsymbol{E}_i \cdot \hat{\boldsymbol{y}},$$
$$\psi_p(x_i, z_i) \equiv \boldsymbol{E}_i \cdot \hat{\boldsymbol{x}}_i. \tag{5.20}$$

Here $\hat{\boldsymbol{y}}$ is taken to point down into the page (for Fig. 5.4) for both coordinate systems. Note that limits of the integrals (5.18, 5.19) actually need to be reduced in magnitude so that $0 \leq \theta_k \leq \pi/2$.

At this point we examine the reflection functions. For TIR, $r_{s/p}(\theta_k)$ is purely a phase, $e^{i\phi_{s/p}(\theta_k)}$. Since the incident beam is narrow, it is reasonable to expand $\phi_{s/p}$ as $\phi_{s/p}(\chi) + \phi'_{s/p}(\chi)(k_x/k_A)$. Inserting this into (5.19) yields

$$\psi_{s/p \text{ refl}}(x_2, z_2) = e^{i\phi_{s/p}(\chi)} \int_{-k_A}^{k_A} f_{s/p}(k_x)\tilde{\psi}(k_x)e^{i[\sqrt{k_A^2-k_x^2}z_2+k_x(x_2+\phi'_{s/p}(\chi)/k_A)]}\,dk_x. \tag{5.21}$$

This expression looks like the expression for the incident beam (5.18) with three changes: 1) the coordinate system has been switched from 1 to 2 simply because the beam has been reflected, 2) there is a phase prefactor due to reflection coefficient at mean angle χ, and 3) the x_2 coordinate has been replaced with $(x_2 - [-\phi'_{s/p}(\chi)/k_A])$. In other words, the reflected beam is the incident beam after a simple $r_{s/p} = 1$ reflection, a multiplication by an overall phase shift, and a parallel shift along x_2 by

$$\Delta_{\text{GH}}^{(A)}(\chi; k) = -\phi'_{s/p}(\chi)/(nk). \tag{5.22}$$

$\Delta_{\text{GH}}^{(A)}$ is Artmann's result for the GH shift. The amazing thing about this formulation is that the exact form of $\tilde{\psi}(k_x)$ does not matter. It could be a fundamental Gaussian, but this is not required. The only requirement is that the incident angle χ be a reasonable "center" or "average" angle for the incident beam.

For a simple dielectric interface reflection, r_s and r_p are given by the Fresnel formulas, written here in form that explicitly uses the critical angle $\chi_c = \arcsin(n_2/n_1)$.

$$r_s(\theta_k) = \frac{\cos\theta_k - [\sin^2\chi_c - \sin^2\theta_k]^{1/2}}{\cos\theta_k + [\sin^2\chi_c - \sin^2\theta_k]^{1/2}},$$
$$r_p(\theta_k) = \frac{-\sin^2\chi_c \cos\theta_k + [\sin^2\chi_c - \sin^2\theta_k]^{1/2}}{\sin^2\chi_c \cos\theta_k + [\sin^2\chi_c - \sin^2\theta_k]^{1/2}}. \tag{5.23}$$

For TIR, $[\sin^2\chi_c - \sin^2\theta_k]^{1/2}$ is positive imaginary and r_s and $(-r_p)$ are of the form

$z/z^* = z^2/|z|^2 = e^{i2\arg z}$ so that the phases are

$$\phi_{\rm s} = -2\arctan\left[\frac{\sqrt{\sin^2\theta_k - \sin^2\chi_{\rm c}}}{\cos\theta_k}\right],$$

$$\phi_{\rm p} = -2\arctan\left[\frac{\sqrt{\sin^2\theta_k - \sin^2\chi_{\rm c}}}{\sin^2\chi_{\rm c}\cos\theta_k}\right] + \pi. \qquad (5.24)$$

The GH shift is calculated straightforwardly from (5.22):

$$\Delta_{\rm GHs}^{(A)} = \frac{\lambda_A}{\pi}\frac{1}{\sqrt{1 - \sin^2(\chi_{\rm c})/\sin^2(\chi)}},$$

$$\Delta_{\rm GHp}^{(A)} = \frac{\Delta_{\rm GHs}}{\sin^2(\chi)/\sin^2(\chi_{\rm c}) - \cos^2(\chi)}. \qquad (5.25)$$

Note that the fact that the GH shifts are different for different polarizations means that a beam containing both s and p polarization experiences a sort of splitting into two adjacent but overlapping beams. As noted before, we will only be concerned with s polarization.

As $(\sin\chi - \sin\chi_{\rm c}) \to 0$, the single-interface GH shifts (5.25) go unphysically to ∞ as $(\sin\chi - \sin\chi_{\rm c})^{-1/2}$. An additional singularity is possible for $\Delta_{\rm GHp}^{(A)}$. This unphysical behavior near the critical angle restricts the use of Artmann's formulas. Unfortunately, we (may) wish to look where the GH shift is near its largest value, since we wish to see whether ray dynamics with the GH shift included can have a significant effect on the decay rates predicted by the sequential tunneling model. The GH shift turns out to be strongest near the critical angle, where we cannot use Artmann's formulas. Also, Artmann's formulas actually predict that the GH shift drops immediately to zero for non-TIR reflection, a behavior that is both intuitively and experimentally incorrect.

There exists a formulation of the GH shift that is far more experimentally accurate than Artmann's near the critical angle.[6] This formulation depends on $\tilde{\psi}(k_x)$, uses a large number of approximations, and involves special functions that are not likely to be in one's favorite math software. Some of the approximations involved break down when the predicted GH shift gets sufficiently large. Because of its length and detail, the calculation is relegated to Appendix E. The next section briefly gives results for

---

[6]There are other formulations. For example, in 1964, Renard [84] derived *different* expressions than Artmann by considering energy flow in the evanescent wave. This result has more problems than Artmann's (see [85]). The formulation we consider now is the most sophisticated to date.

a calculation of the Fresnel kick, using some of the apparatus developed in Appendix E.

5.4 A Calculation of the Fresnel Kick

A beam, incident on a planar interface, that is completely or partially composed of plane waves having θ_k less than χ_c experiences preferential transmission of its low-θ_k components, due to the dramatic dependence of the Fresnel r and t coefficients near the critical angle. This phenomenon has been called Fresnel filtering [60] and can result in a considerable violation of Snell's Law for the transmitted beam, and a smaller (assuming most of the light is reflected) violation of the Law of Reflection, $\theta_{\text{refl}} = \chi$. It is this latter effect that we are interested in; we will refer to the quantity $\theta_{\text{refl}} - \chi$ as the "Fresnel kick".

There are two natural definitions for the magnitude of the Fresnel kick. The result of one definition is given here in detail and the other is only discussed briefly.

5.4.1 Definition 1

One can define θ_{refl} as the mean of $\theta_2 \equiv \arcsin(k_x/k_A)$:

$$\bar{\theta}_2 = \frac{\int |\tilde{\psi}(k_x) r_{\text{s}}(k_x)|^2 \arcsin(k_x/k_A)\, dk_x}{\int |\tilde{\psi}(k_x) r_{\text{s}}(k_x)|^2\, dk_x}. \tag{5.26}$$

We will approximate this integral using calculations similar to those given in Appendix E.

Of course, the asymmetric angular momentum loss of part of the Gaussian beam distorts the beam, causing it to no longer be a Gaussian [86]. We do not worry about this change here.

To calculate $\bar{\theta}_2$ from the first definition, the numerator and denominator are partitioned as $\int_{-\infty}^{-k_A\Delta} + \int_{-k_A\Delta}^{\infty}$. In the second of these integrals, $|r_{\text{s}}|^2 = 1$, which results in integrals which are simply related to the error function. In the first integrals, r_{s} is real, and can be approximated using (E.36). It is consistent to take $\arcsin(k_x/k_A) \approx k_x/k_A$ in both integrals. For the fundamental Gaussian ($m = 0$), Eq. (5.26) becomes

$$\bar{\theta}_2 = \frac{U_1 + \cos^4 \chi_c V_1}{U_0 + \cos^4 \chi_c V_0}, \tag{5.27}$$

where

$$U_i = \int_{-\infty}^{-\Delta} s^i e^{-(k_A\sigma)^2 s^2} N_r^2(s)\, ds,$$

$$V_i = \int_{-\Delta}^{\infty} s^i e^{-(k_A\sigma)^2 s^2}\, ds \qquad (5.28)$$

and $s \equiv k_x/k_A$. The V integrals are simply

$$V_0 = \frac{\sqrt{\pi}}{2k_A\sigma}[1 + \mathrm{erf}(k_A\sigma\Delta)],$$

$$V_1 = \frac{1}{2(k_A\sigma)^2} e^{-(k_A\sigma)^2\Delta^2}. \qquad (5.29)$$

For the U integrals, $N_r^2 \equiv r_{\mathrm{s}}^2 \cos^4 \chi_{\mathrm{c}}$ is approximated as

$$N_r^2 \approx b_1 + b_2\sqrt{-\Delta - s} + b_3 s + b_4 s\sqrt{-\Delta - s}, \qquad (5.30)$$

where

$$b_1 = \cos^4\chi - 6\Delta\cos^2\chi\sin 2\chi + \Delta^2\sin^2 2\chi,$$

$$b_2 = -4\cos\chi\sqrt{\sin 2\chi}(\cos^2\chi - \Delta\sin 2\chi),$$

$$b_3 = -8(\cos^2\chi - \Delta\sin 2\chi)\sin 2\chi,$$

$$b_4 = -4\cos\chi\sqrt{\sin 2\chi}(\cos^2\chi - 2\sin 2\chi - \Delta\sin 2\chi). \qquad (5.31)$$

Using the notation $U_i = \sum_{j=1}^{4} U_{ij}$, the U integrals are constructed from

$$U_{01} = b_1\frac{\sqrt{\pi}}{2k_A\sigma}[1 - \mathrm{erf}(k_A\sigma\Delta)],$$

$$U_{11} = -b_1 V_1,$$

$$U_{03} = -b_3 V_1,$$

$$U_{13} = b_3\left[\frac{\Delta}{2(k_A\sigma)^2}e^{-(k_A\sigma)^2\Delta^2} + \frac{\sqrt{\pi}}{4(k_A\sigma)^3}(1 - \mathrm{erf}(k_A\sigma\Delta))\right],$$

$$U_{02} = b_2 2^{-3/4}\Gamma(3/2)(k_A\sigma)^{-3/2}e^{-(k_A\sigma)^2\Delta^2/2}D_{-3/2}(\sqrt{2}k_A\sigma\Delta),$$

$$U_{12} = -b_2 2^{-5/4}\Gamma(5/2)(k_A\sigma)^{-5/2}e^{-(k_A\sigma)^2\Delta^2/2}D_{-5/2}(\sqrt{2}k_A\sigma\Delta) - \Delta U_{02},$$

$$U_{04} = \frac{b_4}{b_2} U_{12},$$

$$U_{14} = -\frac{b_4}{b_2} \frac{1}{2k_A\sigma} \frac{\mathrm{d}}{\mathrm{d}(k_A\sigma)} U_{02}$$

$$= b_4 2^{-7/4} \Gamma(3/2)(k_A\sigma)^{-7/2} e^{-(k_A\sigma)^2\Delta^2/2}$$

$$\times \left[\frac{3}{2} D_{-3/2}(\sqrt{2}k_A\sigma\Delta) + \sqrt{2}k_A\sigma\Delta D_{-1/2}(\sqrt{2}k_A\sigma\Delta) \right]. \tag{5.32}$$

Some of the integrations have made use of Eqs. (G.7, G.6). Furthermore, the gamma functions are simple constants and $D_{-5/2}$ and $D_{-3/2}$ can be expressed in terms of $D_{-1/2}$ and $D_{1/2}$ via (G.6).

5.4.2 Definition 2

Another definition of θ_{refl} is θ_{peak}, the angular position of the brightest region (in real space) of the beam in the far field limit. This definition corresponds to that of Refs. [80, 60]. The amplitude of the field at a given real-space angular position in the far field is readily found by a stationary phase approximation. We do not give the details here and leave the characterization (including numerical comparison of the two definitions) to future work. In the next chapter we discuss the incorporation of the GH effect into the sequential tunneling model.

Chapter 6

FAILURE AND SUCCESS OF THE GOOS-HÄNCHEN SHIFT APPLIED TO RAY DYNAMICS

This chapter covers two applications of the GH shift, one for dielectric oval resonators, and the other for the dome cavity. In the oval resonator, it is first shown that the sequential tunneling model can significantly underestimate the resonance width when the resonator is non-circular. The next several subsections investigate the necessary technical considerations for applying the GH shift to a circulating mode. The difficulties associated with handling the curved surface, choosing the parameters of the beam, and choosing the starting and stopping critera of the beam are then analysed and solutions are proposed. Finally, in Sec. 6.1.6 we look at the verity of the strong inequalities necessary for the calculation of the GH shift, as given in Appendix E. It is found that one of these inequalities cannot be met. The procedure of bypassing the approximate calculation of the GH shift by numerical integration in k_x is discussed. Our evaluation is that the difficulties and ambiguities associated with the violated inequality, in combination with the difficulties previously discussed, cause the entire GH-augmented sequential tunneling model to become too speculative, and too cumbersome, to merit much further investigation.

On the other hand, application of the GH shift to the dome cavity is found to be an elegant triumph. Here the GH shift is calculated in a simple manner, using the numerical derivative of the plane wave reflection phases for the dielectric mirror. For the stack designs used in Chapter 3, there are no singularities in this calculation

of the GH shift, as there were for the case of TIR at a single interface, described in Chapter 5 and Appendix E. We show that the "V" modes, which were discovered via our dome cavity simulations at non-paraxial geometries, are created and stabilized by the GH shift. The GH shift is clearly *the* mechanism responsible for the existance of the stable V modes, with the phase space map illustrating a saddle-node bifurcation, the birth of a stable-unstable pair of fixed points.

We define here the quantity Δx_{GH} to be the parallel shift along the interface

$$\Delta x_{GH} \equiv \frac{\Delta_{GH}}{\cos \chi}. \tag{6.1}$$

6.1 GH-Augmented Sequential Tunneling

6.1.1 Motivation for a GH-Augmented Sequential Tunneling Model

As can be seen in Fig. 5.2, the sequential tunneling model predicts the resonance widths with reasonable accuracy in a circular billiard. When the the model is applied to a non-circle, however, the resonance widths of WG modes in the TIR regime are significantly underestimated. This discrepancy was first noted in [75]. Due to the negative nature of the conclusion regarding the GH-augmented tunneling model, we only take time to show one specific example of the underestimation of resonance width; the example is typical behavior. We will consider a quadrupole resonator with a deformation of about 10%. The direct or "exact" calculation of the mode is performed numerically by basis expansion and point-matching, using Eqs. (5.2-5.4) with a linear least squares technique. The choice of starting position for the sequential tunneling trajectory, and the averaging of angular momentum for both the wave solution and the classical trajectory, are discussed in Sections 6.1.5 and 6.1.4.

As mentioned in the caption of Fig. 5.1, the quadrupole billiard shape is given in radial coordinates by

$$r(\phi) = r_0[1 + \epsilon \cos(2\phi)]. \tag{6.2}$$

Figure 6.1 shows a WG mode of the $n = 1.5$ quadrupole resonator having a deformation given by $\epsilon = 0.0523226$, which is close to the deformation of the quadrupole used in Fig. 5.1. The value of r_0 is chosen to be $(1 - \epsilon^2/2)^{-1/2} = 0.999316$ so
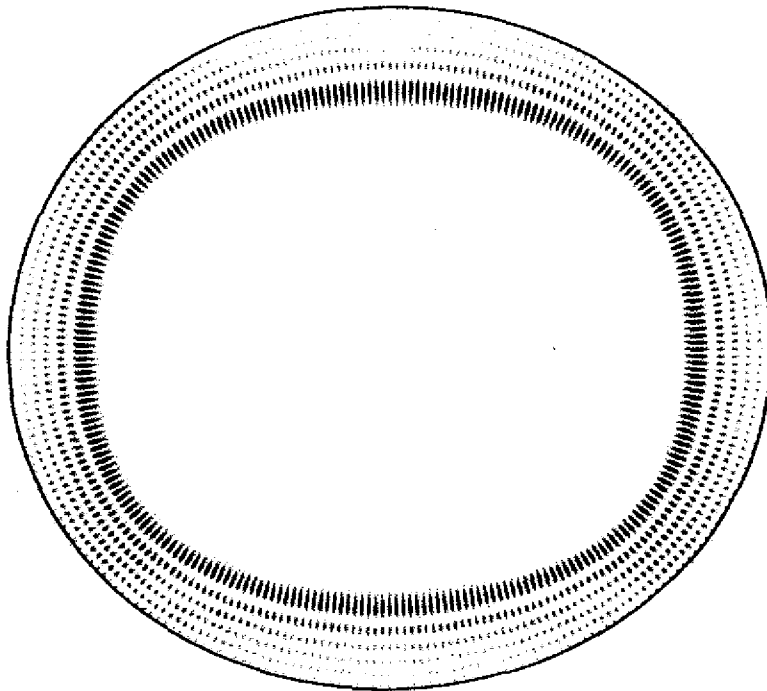
Figure 6.1: Intensity pattern of a whispering gallery mode in the quadrupole.

that the area of the quadrupole is π. The mode calculated has wavenumber $k = 111.480276-0.000462i$ and an integrated average angular momentum of $\langle m \rangle = 129.98$. By scanning the starting $\sin \chi$ value and calculating the average angular momentum of each resulting trajectory, we find trajectories with power-averaged angular momentum values sufficiently close to $\langle m \rangle$. With no GH shift, the matching trajectory yields a predicted value $\mathrm{Im}\, k = -3.25 \times 10^{-5}$, about 14 times smaller in magnitude than the actual value. This trajectory lies on a KAM curve above the four bounce island chain and its surrounding separatrix region.

Section 6.1.2 describes what we believe is the reasonable way to handle the GH shift for a curved surface. When this method is applied to GH shifts in the circle, $\sin \chi$ and angular momentum remain constants of the motion. In other words, the GH effect does basically nothing to affect the classical dynamics in the circle; it increases the bounce-to-bounce increment of s or ϕ, but all other quantities remain unchanged. Given that sequential tunneling produces accurate results for the circle, the GH effect is an ideal candidate for a correction to the sequential tunneling model. For a non-circle, including the GH shift will modify the predicted resonance width as follows.

The GH shift adds a small kick to the position variable, ϕ (or s), of each bounce in phase space. The kicks in ϕ contribute to small changes in $\sin\chi$ upon subsequent iterations of the (non-circle) billiard map. Kicks toward lower $\sin\chi$ cause greater transmission loss, and kicks to larger $\sin\chi$ cause less transmission loss. In the region above the critical angle, transmission decreases exponentially (or faster) with increasing $\sin\chi$. Because of this bias, it can be expected that including the GH effect, which kicks $\sin\chi$ both positively and negatively, should cause a greater average loss rate for the trajectory. Furthermore, the including of the GH effect causes to the billiard map (having position s and momentum $\sin\chi$) to no longer be area-preserving, In such cases one might expect that there would be a diffusion in phase space to lower or greater $\sin\chi$. When the approximate calculation of the GH shift in Appendix E is used, despite the violated inequality discussed in chapter, a diffusion directed toward lower $\sin\chi$ is indeed observed. These questionable results are not presented, other than the statement here that the GH-induced increases in the predicted resonance widths (about a factor of two for the mode given above) are well short of matching the values given by the exact wave calculation.

6.1.2 Implementation of the GH Shift at a Curved Surface

Perhaps the most obvious problem of the GH-augmented sequential tunneling model, although likely one of the least serious, is the issue of how to deal with a curved surface. We use the reasonable solution of Ref. [79] to this problem. Figure 6.2 illustrates the procedure. It is first assumed that $\Delta x_{\text{GH}} \ll R_1$, so that it makes little difference whether the calculated GH shift is wrapped onto the resonator surface or is taken tangent to the resonator surface and then projected onto the surface (we choose to wrap). Second, we make the symmetric choice that the angle of reflection, *relative to the surface at the point of reflection*, be equal to the angle of incidence, relative to the surface at the point of incidence. Finally, for practical ease of calculation, it is assumed, in the calculation of the position of reflection as a function of Δx_{GH}, that the surface can be approximated as a circle over the arc length Δx_{GH}. This latter assumption introduces a small error that is equivalent in effect to having a small error in Δ_{GH}. After the point of reflection is located, the exact normal vector at this position is found, and the symmetry property is strictly enforced.

One result of this GH procedure is that incorporating the GH shift for a perfectly circular billiard only changes the dynamics trivially: the bounce-to-bounce increment
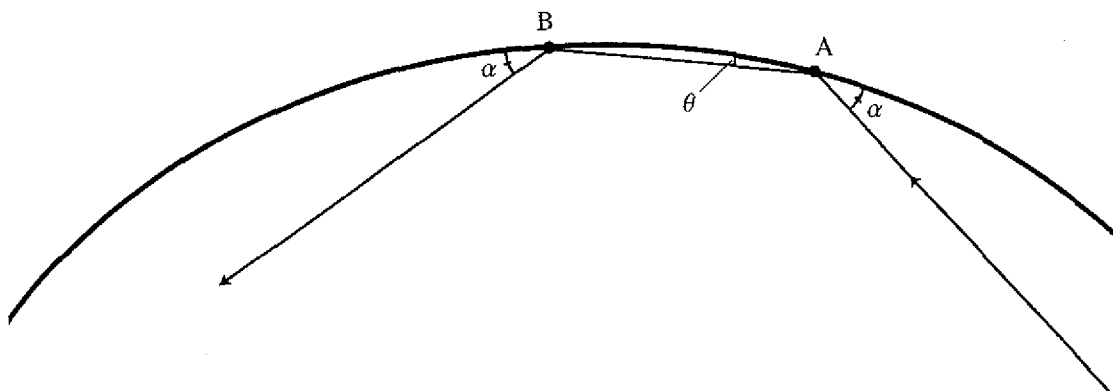
Figure 6.2: Construction of GH shift on a curve. The point of incidence is A, the point of reflection is B, and $\alpha \equiv \pi/2 - \chi$. For numerical simplicity, the arc length AB is not taken to exactly be Δx_{GH}. Instead, B is taken to be the endpoint of the chord shown, where $\theta \equiv \Delta x_{\mathrm{GH}}/(2R_A)$ and R_A is the radius of curvature at A. If the arc is exactly circular, the arc length AB will exactly be Δx_{GH}; if not, the error is negligible.

of ϕ is increased, but the (always constant) value of $\sin \chi$ remains at the value it had without the incorporation of the GH effect. Hence, in a circular billiard, including the GH effect does not alter the prediction of $\operatorname{Im} k$.

6.1.3 Choice of Beam

One of the problems with the use of non-specular sequential tunneling is the assignment of the width and transverse pattern of the traveling beam associated with the mode. As can be clearly seen in the figures of whispering gallery modes in this work (Figs. 5.3, 6.1, 7.3), the whispering gallery modes do not appear to be composed of Gaussian beams (of any transverse order) bouncing around the billiard. Yet this simplification is what is required in a sequential tunneling model. From the visible pattern, it is not clear whether one should use a fundamental Gaussian beam or a Hermite-Gauss beam having the same number of radial peaks as the WG pattern that is being modeled. For simplicity, we use the fundamental Gaussian. This choice is also used in Ref. [79] for the version of the sequential tunneling model discussed at the end of Sec. 5.2.

The choice of the waist radius of the Gaussian beam, w_0, is also a somewhat arbitrary. This parameter enters the formulas for the GH shift given in Appendix E through $\sigma = w_0/\sqrt{2}$. One reasonable choice is that $2w_0$ should be equal to the full

width of the innermost radial antinode of the WG mode. Using the approximations which are given at the end of Sec. 7.3.2, we arrive at

$$
\begin{aligned}
w_0 &\approx \frac{1}{2} \frac{0.79(0.66 - z_1)R}{(n\mathrm{Re}(k)R\sin\chi)^{2/3}} \\
&\approx \frac{1.2R}{(n\mathrm{Re}(k)R\sin\chi)^{2/3}}.
\end{aligned}
\tag{6.3}
$$

where R is either the local or the average/nominal radius of curvature and $z_1 \approx -2.34$ is the first (the least negative) zero of the Airy function approximating the WG mode[1]. This estimation is in general different than the approximation $\sigma \approx R(n\mathrm{Re}(k)R)^{-2/3}$ used by Hentschel and Schomerus in Ref. [79].

One way to check the reasonablity of our choice of w_0 is to consider the n-bounce islands of the quadrupole SOS which bracket the desired region of $\sin\chi$. For $n = 1.5$, the region of most practical interest is between the four bounce and the five bounce islands shown in Fig. 5.1. For sufficiently large k, one can numerically find the modes of transverse order zero which live in these islands and numerically measure w_0. Alternatively, w_0 for these modes can be estimated by the Gaussian ABCD method [34].

6.1.4 Choice of Stopping Criterion

A non-specular correction to the classical dynamics can cause the several-cycle average of $\sin\chi$ to slowly vary with time, so that the cycle-averaged power loss rate of the trajectory is not constant. When the GH shift is turned off, there is no difficulty (or little difficulty[2]) in assigning a single effective value of $\mathrm{Im}\,k$ to the exponential decay behavior, as done in Sec. 5.2. However, to extract a single value for the resonance width when the loss rate is increasing requires that a procedure for extracting the average loss rate be established. The most natural definition, perhaps, is to integrate the instantaneous loss rate over time, weighted by the instantaneous power. For a

---

[1]This approximation is derived for metallic resonators, but this is the best estimate available.

[2]For specular reflection, diffusion of $\sin\chi$ may still exist within chaotic regions, which are bounded by KAM surfaces/curves [75]. Once the chaotic region is saturated however, there is no further diffusion. Additionally, there is no diffusion for KAM trajectories.

trajectory running from $t = 0$ to $t = T_f$ this yields an average loss rate of

$$\bar{R} = \frac{\int_0^{T_f} R(t)P(t)\,dt}{\int_0^{T_f} P(t)\,dt}$$
$$= \frac{P(0) - P(T_f)}{\int_0^{T_f} P(t)\,dt} \approx \frac{P(0)}{\int_0^{\infty} P(t)\,dt}, \tag{6.4}$$

where P is the power associated with the ray/beam. In this protocol, a precisely defined stopping time T_f is not necessary, and one may truncate the trajectory after further integration will negligibly effect the rate. The predicted value of Im k is simply

$$\mathrm{Im}\,k = -\frac{\bar{R}}{2c}. \tag{6.5}$$

There is another possible complication to the issue of stopping criterion. For a quantum particle momentarily confined in a classically chaotic billiard which has physical openings as escape channels, it is found [87] that the quantum probability for the particle to remain in the billiard, $P(t)$, decreases as predicted[3] by classical dynamics (exponentially/diffusively), for times less than the Heisenberg time, $t_H = \hbar/\Delta$, where Δ is the mean energy level spacing. For times much greater than t_H, $P(t)$ obeys a more complicated decay law instead of the classically expected diffusion. The question arises as to whether t_H represents a loose upper limit on the time that we can consider the classical trajectories to be of value. It is not clear whether the situation described above is relevant to the current goal of modeling a single quasimode with a trajectory. Because we are not considering the ray to represent a temporal wavepacket, it seems likely that t_H is not important to our model. For the cases we consider, the timescale on which the decay rate changes significantly is generally much greater than t_H, indicating that the effects of the GH shift would be negligible if the cutoff were to be imposed.

6.1.5 Choice of Initial Condition

The choices of initial and stopping conditions are difficulties in the following sense. We are attempting to model a mode which is a stationary state (or at least a steady state) by a trajectory that starts somewhere at some time, diffuses in phase space, and decays to oblivion. We must imagine such trajectories to be continuously created

---

[3]There are corrections, which are the subject of Ref. [88].

in time, perhaps at a continuum of ϕ values. In picking the starting conditions, we are not trying to match the decay rates of the trajectories to the decay rate of the quasimode—that correspondence is what we are testing. We are, however, attempting to match other quantities such as the average angular momentum. One has the choice of starting the trajectory so that either its initial few cycles or its power-weighted-average matches the average angular momentum of the mode. It seems that the latter is the best choice, and we attempt to set this correspondence. Note that we can only choose to match one property, as there is only one free parameter, namely $\sin \chi$, in the ϕ-symmetrized family of trajectory starting positions; for instance, we can match average angular momentum but not its variance. There is further ambiguity in how one calculates the average angular momentum of a directly calculated mode. On can take the average to be $\sum m|a_m|^2 / \sum |a_m|^2$ or $\sum m|b_m|^2 / \sum |b_m|^2$ or perhaps $\sum m(|a_m|^2+|b_m|^2)/ \sum(|a_m|^2+|b_m|^2)$. Alternatively, one can numerically integrate the physical angular momentum density over a region A, taking A to be the interior of the resonator, or all space, or perhaps the interior (dielectric region) plus a region of the exterior several wavelengths wide surrounding the interior (to include the evanescent wave but not a significant fraction of the escaping wave). After this integration one divides by $(1/\omega)$ times the integrated energy density to obtain the average m (see Appendix F). We routinely compare methods of calculating $\langle m \rangle$. As an example, for the ellipse of area π, eccentricity 0.5, and refractive index $n = 1.2$, the mode with $k = 92.055 - 0.0017i$ had the following averages $\langle m \rangle$

$$103.57 : \qquad \sum m|a_m|^2 / \sum |a_m|^2$$
$$105.85 : \qquad \sum m|b_m|^2 / \sum |b_m|^2$$

$$102.90 : \qquad \text{numerical integration, } A \text{ is interior}$$
$$102.91 : \qquad \text{num. int., } A \text{ is interior} + 2\lambda\text{-wide region of exterior} \qquad (6.6)$$

As we are comparing a trajectory inside the billiard with the resonator mode, it appears that the best definition to use is the numerical integration over the angular momentum density. The agreement between the last two values above indicates that it is sufficient to integrate over the interior of the billiard. In the comparisons given in Sec. 6.1.1, we calculate $\langle m \rangle$ by numerical integration over the resonator interior.

We then pick a trajectory that has a power-weighted average angular momentum that matches $\langle m \rangle$.

Finally, we note that the GH shifts themselves should be included in the time/distance traversed. The power associated with the time spent in a GH shift, $n\Delta x_{\mathrm{GH}}/c$, is taken to be the average of the power values before and after the complete reflection event. The angular momentum associated with this time period is also taken to be the before-and-after average.

6.1.6 The Violated Inequality

To examine the strong inequalities necessary for the calculation of the GH shift performed in Appendix E, we reconsider the specific resonator and mode used in Sec. 6.1.1.

Calculating the GH shift according to Appendix E, and following the procedures determined in Secs. 6.1.2-6.1.5, one finds that the average trajectory local angular momentum is manifestly non-monotonic in the initial value of $\sin\chi$, meaning that a range of trajectories must be considered. For trajectories begining at $\phi = 0$, the initial values of $\sin\chi$ range from 0.7712 to 0.7715. Inspection of the GH trajectories shows a slow downward diffusion in $\sin\chi$, and an accompanying increase in the loss rate. While the power of the trajectory rays are greater than 10%, the GH trajectories lie above the four-bounce island, with $\sin\chi$ between about 0.81 and 0.72. This ensures that the requirement given in Eq. (E.35) is well met. The value of $(kn\sigma)^{-1}$ typically lies between 0.17 and 0.19, with σ between about 0.032 and 0.034. This marginally satisfies the principal requirement of the GH calculation that $(kn\sigma)^{-1}$ be small (E.8). The quantity Δ, defined in Eq. (E.2), ranges from about 0.07 to 0.22, and cannot be considered large, justifying the necessity of the complicated, σ-dependent formulation of Δ_{GH} instead of the simple Artmann formulation. The inequality in Eq. (E.29), however, is violated, with the $z_2 \approx 0.7$ being somewhat greater than $2kn\sigma^2 \approx 0.4$, instead of being much less.

Using Eq. (6.3) with $R = 1$, we have the estimate

$$kn\sigma^2 \approx 0.7(nk)^{-1/3}(\sin\chi)^{-4/3}. \tag{6.7}$$

Decreasing nk to increase the RHS will cause trouble for the other approximations. Decreasing $\sin\chi$ enough to make the inequality (E.29) true by a factor of 3 or 4 puts us in a region of phase space that is undesirable because it has very shape-specific properties and represents TIR modes only for a very low index of refraction n. On the

other hand, we could attempt to decrease z_2 by going to higher angular momentum modes, but it is clear this idea has multiple problems. Use of the Fresnel kick seems inappropriate, as we are not in the limit of very large z_2; z_2 necessarily lies in an intermediate regime. For all oval dielectric problem of interest, the strong inequality is not met. This indicates that a z_2-dependent nonspecular reflection calculation, via numerical integration of Eq. (E.1), is necessary.

At this point, because of its limited regimes of applicability and its complexity of calculation, it appears that inclusion of the GH shift into dynamics is not a practical way to calculate properties of modes. However, if a corrected version of our procedure using numerical integration of $S(k_x)$ were to accurately estimate mode quantities, it would aid understanding of the ray-wave transition. Unfortunately, the dependence on z_2, although in principle numerically calculable, introduces additional difficulties or ill-guided choices that must be made. A dependence on z_2 means that the relected beam, or its position of maximum intensity, curves in space. To take this curve to be "the ray", including its in-flight bending toward more normal reflection, appears to be an additional speculative feature that stretches the already troubled GH-augmented sequential tunneling model beyond reasonable believability. For this reason we have chosen not to implement the numerically intensive determination of the curved ray and its intersection with the billiard.

We next present a situation in which the GH effect accurately predicts mode phenomena.

6.2   The "V" Mode in the Dome Cavity

Figure 6.3 shows a contrast-enhanced y-z slice of the E_x field of a peculiar V-shaped $m = 1$ mode in the dome cavity. Azimuthally, the electric field strength decays to zero at $\phi = 0, \pi$. The inset shows the plot of the s-polarized plane wave polar angle distribution, $\sin\theta_k |S_d(\theta_k)|^2$. Using this distribution, the mean polar angle, $\langle \theta_k \rangle$, is found by rough integration to be $53.8°$. The distribution for p-polarized light is also plotted, but is indistinguishable from the x axis; the mode is unequivocally s-polarized.

Our early investigations into nonparaxial modes in the dome cavity revealed V modes [24], although it was not until our later investigations into the GH shift that Jens Nöckel suggested that the GH shift might explain these modes. Subsequent investigations into the phenomena have revealed that the GH modes can be s or p

Ex
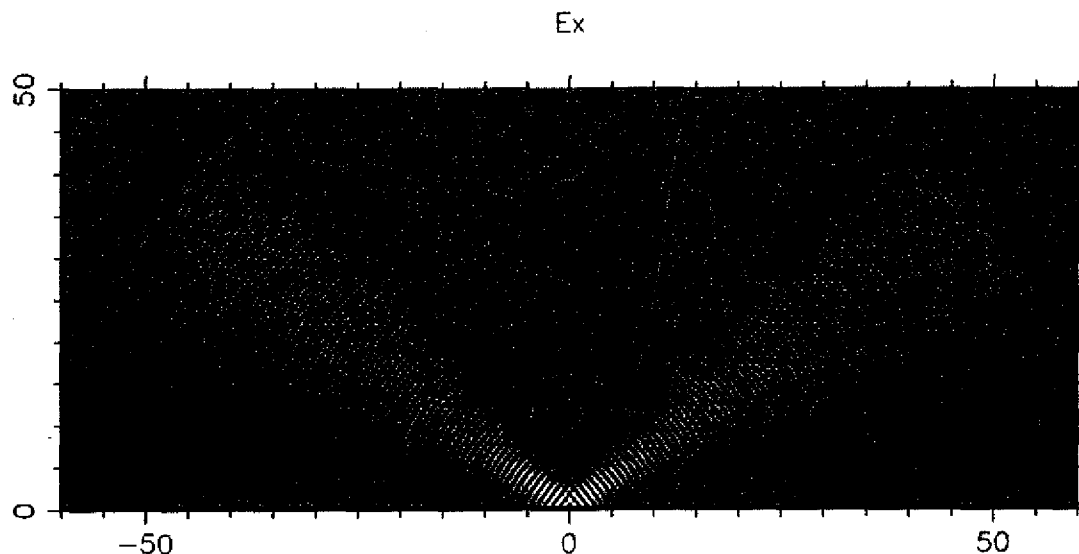
Figure 6.3: An s-polarized V mode. The inset shows the distribution of plane wave components versus θ_k in degrees. The mode/cavity parameters are $m = 1$, $R = 60$ (M2 spherical), $z_1 = 0.4$, $k = 8.31728 - 5.56 \times 10^{-6}i$ and the Bragg mirror has a 1λ front spacer layer ($n = 3.52$) followed by 22 quarter-wave layer pairs of indices 3.0 and 3.52.

polarized, and are not a nonparaxial version of the paraxial Gaussian modes: both the V modes and the low order Gaussian modes coexist.[4] The GH modes also exist in the scalar version of the cavity problem, as long as there is a Bragg mirror. When M1 is conducting, the V modes have not been found, and, by the theory we discuss next, are not predicted to exist.

As previously noted, a Bragg mirror general has phase shifts ϕ_s and ϕ_p which are functions of θ_k. In Chapter 5 it was shown that the simple GH shift is proportional to $d\phi/d\theta_k$. Using Eq. (5.22), we see that in the dome cavity

$$\Delta x_{GH} = -\frac{1}{k \cos \theta_k} \frac{d\phi_{s/p}(\theta_k; k)}{d\theta_k}. \tag{6.8}$$

---

[4]The coexistence was discovered late, and in Ref. [24] it is misleadingly stated that the fundamental Gaussian, in a cavity with a Bragg mirror, can split and turn into a V mode as the cavity becomes less paraxial. What actually occurs is that, as the fundamental Gaussian is numerically followed from paraxial to non-paraxial, it undergoes an anticrossing with the V mode and thus "becomes" the V mode. The important distinction is that there are two modes involved in this metamorphosis, not one. We have seen this anticrossing with the fundamental Gaussian occur with both s and p polarized V modes.
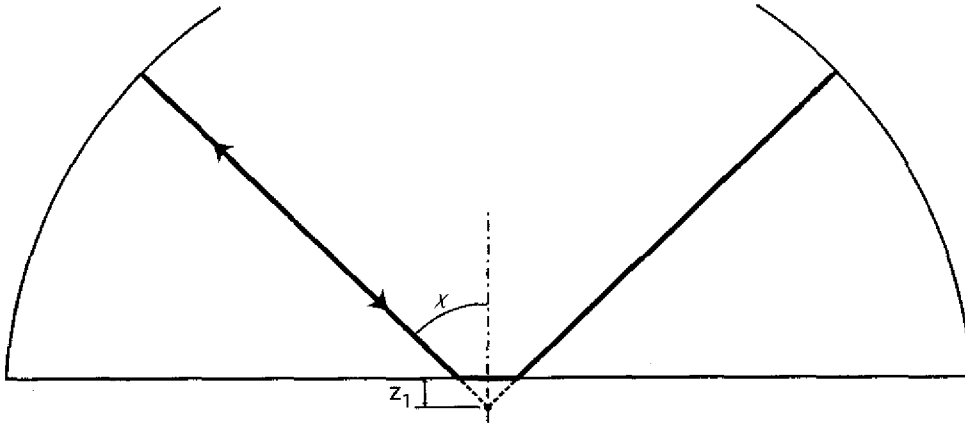
Figure 6.4: Periodic orbit for a V mode.

If we assume the V modes are associated with the simple self-retracing bouncing ball orbit which has angle of incidence χ at M1 and is normal at M2, simple geometry leads to the prediction that a shift along the surface of M1 of length $2z_1 \tan \chi$ is necessary (Fig. 6.4). Supposing the GH effect to account for this shift, one sees that there is a GH-predicted value of χ given k and z_1; the predicted value satisfies

$$-\frac{\mathrm{d}\phi_{\mathrm{s/p}}(\theta_k; k)}{\mathrm{d}\theta_k}\bigg|_{\theta_k=\chi_{\mathrm{GH}}} = 2kz_1 \sin \chi_{\mathrm{GH}}. \qquad (6.9)$$

The choice of whether to use ϕ_{s} or ϕ_{p} is put in by hand (if one is trying to model an existing s-polarized mode one uses ϕ_{s}). For the mode shown in Fig. 6.3, $\chi_{\mathrm{GH}} = 54.0°$, which agrees extremely well with the mean numerical polar angle, $\langle\theta_k\rangle = 53.8°$. For a p-polarized V mode with the same cavity parameters, $k = 8.31484 - 1.65 \times 10^{-5}i$ and $\langle\theta_k\rangle = 34.5°$, the predicted value is $\chi_{\mathrm{GH}} = 38.0°$. We have also found what appear to be higher order V modes which have more than one peak in θ_k. At the same cavity parameters as the modes above, there is a second order V mode at $k = 8.31543 - 6.2 \times 10^{-6}i$ with mean angle $\langle\theta_k\rangle = 52.0°$. For this mode, $\chi_{\mathrm{GH}} = 55.2°$. We conclude that the self-retracing periodic orbits predicted by the GH shift provide a compelling description of the V modes.

One can numerically calculate the full round trip SOS-like map, $T : (p_{\theta,i}, \theta_i) \mapsto (p_{\theta,i+1}, \theta_{i+1})$, for the two dimensional dome billiard with GH included. The position and momentum coordinates we use are θ, the polar angle of the position in real space a bouncing particle, and $p_\theta = x \times \hat{v}$, where $\hat{v}$ is the unit velocity. The SOS points

$(p_{\theta,i}, \theta_i)$ are taken when $r = R$ (when the particle is at the M2 surface), and when the particle is on the right hand side of the billiard. At $(p_\theta, \theta) = (0, \chi_{GH})$, T has a fixed point. By looking at the monodromy matrix, the derivative of T given by

$$DT \equiv \begin{pmatrix} \frac{\partial p_{\theta,i+1}}{\partial p_{\theta,i}} & \frac{\partial p_{\theta,i+1}}{\partial \theta_i} \\ \frac{\partial \theta_{i+1}}{\partial p_{\theta,i}} & \frac{\partial \theta_{i+1}}{\partial \theta_i} \end{pmatrix}, \qquad (6.10)$$

we can determine whether the fixed point is stable or unstable, and whether the map is area preserving or not near the fixed point. We performed a rough calculation of DT for the orbit corresponding to the V mode of Fig. 6.3. The eigenvalues of this DT are a complex conjugate pair, and they appear to very close to the unit circle. If the pair were exactly on the unit circle, the situation would correspond to a stable fixed point of an area preserving map (e.g. [72]). There is no reason to believe that T is truly area preserving (true when the determinant of DT is unity), and it is expected that a careful calculation will show that the deviation of $\det(DT)$ from 1 is larger than can be accounted for by numerical round-off error. In general, $\det(DT)$ may be expected to vary with position in phase space. The certain result is that an orbit started near the periodic orbit remains in the vicinity of the orbit for a large number of bounces.

The V mode lies in the stable fixed point that is part of a stable-unstable pair of fixed points that comes into being as z_1 is varied accross a special value z_c. (This creation is called a saddle-node bifurcation.) For our parameters, Eq. (6.9) goes from having only one trivial solution at $\chi_{GH} = 0$ at $z_1 > z_c$ to having two solutions at $z_1 = z_c$ to having three solutions as z_1 becomes less than z_c. Further discussion of the V mode and its classical origin is expected to be published in Optics Letters.

Chapter 7

THE GENERALIZED BORN-OPPENHEIMER APPROXIMATION FOR WG MODES

In this chapter we wish to investigate the application of the generalized Born-Oppenheimer approximation (BOA) to WG modes in dielectric oval resonators and to WG-type waves traveling along a curved dielectric interface. The BOA approximation is also known as the adiabatic approximation, but this appellation is less specific. The application of the BOA to WG modes/waves in dielectric resonators involves an extension of the work of Zaitsev *et al.* [89, 90], which implemented the BOA for closed cavities with Dirichlet (conducting) boundary conditions. The first work on the subject was done by Lord Rayleigh, who coined the term "whispering gallery wave".[1]

Before implementing the BOA, we will "guess" the approximate solution to the problem of a WG-type wave traveling along a curved dielectric boundary, *for which the radius of curvature, R, varies slowly with path length, s.*[2] We assume that the wave can be approximately described by a factored wavefunction $\Psi(\rho, s) = \Phi(\rho)\psi(s)$ where s measures distance along the boundary and ρ measures distance away from

---

[1] Or something like this.

[2] Speaking of guesses, it seems a reasonable guess that the type of approximation we are making should be valid in the limit $k \to \infty$ even if the strictly geometric quantity dR/ds is of order one. As will be shown later, this turns out to be false.

the boundary. A wave is launched at $s = 0$ and propagates a distance s along the boundary. One might guess, by extension of the well-known solution to the circle resonator, that there is an effective Helmholtz-like equation for this "azimuthal" propagation,

$$\psi'' + U^2(s)\psi = 0. \tag{7.1}$$

(In the case of a circular resonator of radius R, the separation of variables for Ψ would be exact, and one would have $\rho \equiv R - r$, $s \equiv R\phi$, and $U = \text{const.} = m/R$, where m is an integer and is associated with the angular momentum.)

To find more about U, we note that in the dielectric circle, $m = UR$ satisfies the resonance condition (5.8). (In the conducting circle, m satisfies $J_m(nkR) = 0$.) Thus we could imagine that the approximate solution for our WG traveling wave (having in this context a real wavenumber $k = \omega/c$), has $U(s)$ satisfying a local resonance condition:

$$J_{R(s)U(s)}(nkR(s))H^{(1)}_{R(s)U(s)-1}(kR(s)) = nJ_{R(s)U(s)-1}(nkR(s))H^{(1)}_{R(s)U(s)}(kR(s)). \tag{7.2}$$

The reasonable guesses so far will be justified by the use of the BOA.

Solutions of interest to the resonance equation (7.2) for whispering gallery modes typically have

$$\text{Re}(k)R(s) < \text{Re}\,[U(s)]R(s) < n\text{Re}\,(k)R(s). \tag{7.3}$$

The middle quantity can be thought of as a local angular momentum, measured from the local center of curvature. The justification for the second inequality comes from the fact that the classical angular momentum of a ray traveling tangentially to the boundary at s, is $n\text{Re}(k)R(s)$. This is the maximum angular momentum a ray inside the resonator/billiard can have, and we expect the actual WG modes to have a smaller local angular momentum. The first inequality comes from the fact that we are usually interested in trajectories above the critical angle; trajectories that have some bounces below the critical angle are short lived. The angular momentum of a ray is $n\text{Re}(k)R(s)\sin\chi$, and equating this with $\text{Re}(k)R(s)$ means that the ray is at the critical angle. It is for the above reasons we can say that $U = O(nk)$ (or $U = O(k)$); n is always assumed to be $O(1)$.

After the performing the BOA, the next step essentially involves the WKB ap-

proximation, (which was used in optics by Lord Rayleigh some time before Wentzel, Kramers, or Brillouin applied it the Schrödinger equation). Since $R(s)$ varies slowly, the $U(s)$ that solves (7.2) can be expected to vary slowly. Thus it is reasonable to expect that the forward-traveling solution to (7.1) is approximately given by the WKB solution

$$\psi^{(+)}(s) \propto |U(s)|^{-1/2} \exp(i \int^{s} U(s')\,\mathrm{d}s').$$

$$(7.4)$$

If k is constrained to be real, it turns out that solutions to Eq. (7.2) have $\mathrm{Im}\, U(s) > 0$, which correspond to the expected losses in ψ due to radiation. For a wave traveling from s_1 to s_2, this leads to a simple approximate power transmittance of

$$T_{12} \approx \exp\left(-2 \,\mathrm{Im} \int_{s_1}^{s_2} U(s')\,\mathrm{d}s'\right).$$

$$(7.5)$$

Finally, one may ask, "What is the relation between the resonance width $2|\mathrm{Im}\, k|$ for a resonator mode and the integrated average of $\mathrm{Im}\, U(s)$ (calculated with k constrained to be real)?" Is it simply $\langle \mathrm{Im}\, U \rangle = n|\mathrm{Im}\, k|$? Is it perhaps given by $R\langle \mathrm{Im}\, U \rangle / |n\mathrm{Im}\, k| = R\langle \mathrm{Re}\, U \rangle / (n\mathrm{Re}\, k)$, where the latter quantity is essentially the ratio of angular momentum to linear momentum? We will attempt to answer these questions after deriving and justifying the BOA and the WKB approximations. We will apply our results to the dielectric ellipse, a wave problem which is believed to be non-separable [75].[3]

7.1 Boundary Coordinates and the Helmholtz Equation

When working with WG modes it is useful to use a coordinate system (ζ, ξ) for which one differential vector, $\mathrm{d}\xi$, lies tangent to the resonator boundary δG, and the other, $\mathrm{d}\zeta$, is perpendicular to the resonator boundary. If parameters such as k are chosen so that the wavefunctions of interest decay rapidly enough in the direction of the interior of the resonator, we should not have to worry about the fact that a point on the plane (x, y) corresponds to multiple pairs (ζ, ξ). For points near δG, it is obvious which values of ζ and ξ are correct. We will use the specific boundary coordinates (ρ, s),

---

[3]The conducting ellipse problem is separable in elliptic coordinates.

where ρ increases as one moves toward the *interior* of the resonator and s measures distance along δG, from some preordained point where $s = 0$.

A curve such as δG is naturally parameterized by the function $R(s)$ where R is the instantaneous radius of curvature. We will sometimes use the curvature $\kappa(s) \equiv 1/R(s)$. Certain initial conditions and conventions must be set to obtain the parametrization map $s \mapsto (x_G(s), y_G(s))$. The conventions we use are 1) the resonator shape is convex ($R > 0$ everywhere), 2) we place the origin of our Cartesian coordinate system inside the resonator, 3) $y_G(0) = 0$, $x_0 \equiv x_G(0) > 0$, and $dx_G(0)/ds = 0$ (this constrains the position of the origin), and 4) s advances positively in the ccw sense.

The parametrization map for points on the boundary ($\rho = 0$) is

$$x_G(s) = x_0 - \int_0^s \sin c(s')ds',$$

$$y_G(s) = \int_0^s \cos c(s')ds', \tag{7.6}$$

where

$$c(s) \equiv \int_0^s \kappa(s')ds'. \tag{7.7}$$

(Note that $c(s) + \pi/2$ is the angle of the curve's tangent at s.) The coordinate transformation for a general point (ρ, s) is

$$x = x_0 - \int_0^s \sin c(s')ds' - \rho \cos c(s)$$

$$y = \int_0^s \cos c(s')ds' - \rho \sin c(s). \tag{7.8}$$

There is typically no analytical inversion of this transformation.

Our next goal is to write the 2D Helmholtz equation, $[\partial^2/\partial x^2 + \partial^2/\partial y^2 + n^2 k^2]\Psi = 0$, in the (ρ, s) coordinate system. The result is Eq. (7.20), which is given in the book by Babic and Buldyrev [91]. We derive this equation as follows.

First, we note that $(\partial/\partial x)_y$ and $(\partial/\partial y)_x$, when acting on an expression in (ρ, s)

coordinates, are

$$\frac{\partial}{\partial x} = \left(\frac{\partial s}{\partial x}\right)_y \frac{\partial}{\partial s} + \left(\frac{\partial \rho}{\partial x}\right)_y \frac{\partial}{\partial \rho}$$

$$\frac{\partial}{\partial y} = \left(\frac{\partial s}{\partial y}\right)_x \frac{\partial}{\partial s} + \left(\frac{\partial \rho}{\partial y}\right)_x \frac{\partial}{\partial \rho}. \tag{7.9}$$

Squaring yields

$$\frac{\partial^2}{\partial x^2} = \left(\frac{\partial s}{\partial x}\right)_y^2 \frac{\partial^2}{\partial s^2} + \left(\frac{\partial \rho}{\partial x}\right)_y^2 \frac{\partial^2}{\partial \rho^2} + 2\left(\frac{\partial s}{\partial x}\right)_y \left(\frac{\partial \rho}{\partial x}\right)_y \frac{\partial^2}{\partial \rho \partial s}$$

$$+ \left[\left(\frac{\partial s}{\partial x}\right)_y \left(\frac{\partial}{\partial s}\left(\frac{\partial s}{\partial x}\right)_y\right)_\rho + \left(\frac{\partial \rho}{\partial x}\right)_y \left(\frac{\partial}{\partial \rho}\left(\frac{\partial s}{\partial x}\right)_y\right)_s\right] \frac{\partial}{\partial s} +$$

$$+ \left[\left(\frac{\partial s}{\partial x}\right)_y \left(\frac{\partial}{\partial s}\left(\frac{\partial \rho}{\partial x}\right)_y\right)_\rho + \left(\frac{\partial \rho}{\partial x}\right)_y \left(\frac{\partial}{\partial \rho}\left(\frac{\partial \rho}{\partial x}\right)_y\right)_s\right] \frac{\partial}{\partial \rho}$$

$$\frac{\partial^2}{\partial y^2} = \left(\frac{\partial s}{\partial y}\right)_x^2 \frac{\partial^2}{\partial s^2} + \left(\frac{\partial \rho}{\partial y}\right)_x^2 \frac{\partial^2}{\partial \rho^2} + 2\left(\frac{\partial s}{\partial y}\right)_x \left(\frac{\partial \rho}{\partial y}\right)_x \frac{\partial^2}{\partial \rho \partial s}$$

$$+ \left[\left(\frac{\partial s}{\partial y}\right)_x \left(\frac{\partial}{\partial s}\left(\frac{\partial s}{\partial y}\right)_x\right)_\rho + \left(\frac{\partial \rho}{\partial y}\right)_x \left(\frac{\partial}{\partial \rho}\left(\frac{\partial s}{\partial y}\right)_x\right)_s\right] \frac{\partial}{\partial s} +$$

$$+ \left[\left(\frac{\partial s}{\partial y}\right)_x \left(\frac{\partial}{\partial s}\left(\frac{\partial \rho}{\partial y}\right)_x\right)_\rho + \left(\frac{\partial \rho}{\partial y}\right)_x \left(\frac{\partial}{\partial \rho}\left(\frac{\partial \rho}{\partial y}\right)_x\right)_s\right] \frac{\partial}{\partial \rho} \tag{7.10}$$

These two equations are identical, except that x and y have been interchanged.

Now we detour to solve (7.8) for ρ:

$$\rho(s,x) = \frac{-x + x_0 - \int_0^s \sin c(s')ds'}{\cos c(s)}$$

$$\rho(s,y) = \frac{-y + \int_0^s \cos c(s')ds'}{\sin c(s)}. \tag{7.11}$$

These half-inverted expressions, combined with $x(\rho, s)$ and $y(\rho, s)$ from (7.8), are all we need to obtain all of the derivative quantities in (7.10) if we employ some rules of partial differentiation when a coordinate transformation $(x, y) \leftrightarrow (a, b)$ is made. One such rule is

$$\left(\frac{\partial x}{\partial a}\right)_y = \left(\frac{\partial x}{\partial a}\right)_b + \left(\frac{\partial x}{\partial b}\right)_a \left(\frac{\partial b}{\partial a}\right)_y. \tag{7.12}$$

Another such rule is

$$\left(\frac{\partial a}{\partial x}\right)_y = \frac{1}{\left(\frac{\partial x}{\partial a}\right)_y}.$$

(7.13)

The derivatives that we need (expressed as functions of ρ, s) are

$$\left(\frac{\partial s}{\partial x}\right)_y, \qquad \left(\frac{\partial s}{\partial y}\right)_x, \qquad \left(\frac{\partial \rho}{\partial x}\right)_y, \qquad \left(\frac{\partial \rho}{\partial y}\right)_x.$$

So

$$\left(\frac{\partial \rho}{\partial x}\right)_y = \frac{1}{\left(\frac{\partial x}{\partial \rho}\right)_y}$$

$$= \frac{1}{\left(\frac{\partial x}{\partial \rho}\right)_s + \left(\frac{\partial x}{\partial s}\right)_\rho \left(\frac{\partial s}{\partial \rho}\right)_y}$$

$$= \frac{1}{\left(\frac{\partial x}{\partial \rho}\right)_s + \left(\frac{\partial x}{\partial s}\right)_\rho / \left(\frac{\partial \rho}{\partial s}\right)_y}$$

$$= \left[-\cos c(s) + \left(-\sin c(s) + \rho \kappa(s) \sin c(s) \right) \right.$$

$$\left. \times \left(\frac{-\kappa(s)\cos c(s)}{\sin^2 c(s)} \underbrace{\left(-y + \int_0^s \cos c(s') ds' \right)}_{\rho \sin c(s)} + \frac{\cos c(s)}{\sin c(s)} \right)^{-1} \right]^{-1}$$

$$= -\cos c(s),$$

(7.14)

where the last step involves some algebra. Similarly

$$\left(\frac{\partial \rho}{\partial y}\right)_x = \frac{1}{\left(\frac{\partial y}{\partial \rho}\right)_s + \left(\frac{\partial y}{\partial s}\right)_\rho / \left(\frac{\partial \rho}{\partial s}\right)_x}$$

$$= -\sin c(s),$$

(7.15)

$$\left(\frac{\partial s}{\partial x}\right)_y = \frac{1}{\left(\frac{\partial x}{\partial s}\right)_\rho + \left(\frac{\partial x}{\partial \rho}\right)_s \left(\frac{\partial \rho}{\partial s}\right)_y}$$

$$= \frac{-\sin c(s)}{1 - \rho \kappa(s)},$$

(7.16)

$$\left(\frac{\partial s}{\partial y}\right)_x = \frac{1}{\left(\frac{\partial y}{\partial s}\right)_\rho + \left(\frac{\partial y}{\partial \rho}\right)_s \left(\frac{\partial \rho}{\partial s}\right)_x}$$

$$= \frac{\cos c(s)}{1 - \rho \kappa(s)}. \tag{7.17}$$

Substituting back into (7.10) yields

$$\frac{\partial^2}{\partial x^2} = \frac{\sin^2 c(s)}{(1 - \rho\kappa(s))^2}\frac{\partial^2}{\partial s^2} + \cos^2 c(s)\frac{\partial^2}{\partial \rho^2}$$
$$+ 2\cos c(s)\left(\frac{\sin c(s)}{1 - \rho\kappa(s)}\right)\frac{\partial^2}{\partial \rho \partial s} - \kappa(s)\sin c(s)\left(\frac{\sin c(s)}{1 - \rho\kappa(s)}\right)\frac{\partial}{\partial \rho}$$
$$+ \left(\frac{\sin c(s)}{1 - \rho\kappa(s)}\right)\left(\frac{2\kappa(s)\cos c(s)}{1 - \rho\kappa(s)} + \frac{\rho\kappa'(s)\sin c(s)}{(1 - \rho\kappa(s))^2}\right)\frac{\partial}{\partial s}, \tag{7.18}$$

$$\frac{\partial^2}{\partial y^2} = \frac{\cos^2 c(s)}{(1 - \rho\kappa(s))^2}\frac{\partial^2}{\partial s^2} + \sin^2 c(s)\frac{\partial^2}{\partial \rho^2}$$
$$- 2\sin c(s)\left(\frac{\cos c(s)}{1 - \rho\kappa(s)}\right)\frac{\partial^2}{\partial \rho \partial s} - \kappa(s)\cos c(s)\left(\frac{\cos c(s)}{1 - \rho\kappa(s)}\right)\frac{\partial}{\partial \rho}$$
$$+ \left(\frac{\cos c(s)}{1 - \rho\kappa(s)}\right)\left(\frac{-2\kappa(s)\sin c(s)}{1 - \rho\kappa(s)} + \frac{\rho\kappa'(s)\cos c(s)}{(1 - \rho\kappa(s))^2}\right)\frac{\partial}{\partial s}. \tag{7.19}$$

Thus we have for the Helmholtz equation:

$$\left[\frac{\partial^2}{\partial \rho^2} + \frac{-\kappa(s)}{1 - \rho\kappa(s)}\frac{\partial}{\partial \rho} + \frac{1}{(1 - \rho\kappa(s))^2}\frac{\partial^2}{\partial s^2} + \frac{\rho\kappa'(s)}{(1 - \rho\kappa(s))^3}\frac{\partial}{\partial s} + n^2 k^2\right]\Psi = 0. \tag{7.20}$$

This equations checks with the result of Ref. [91].

7.2 Separation of Variables via the BOA

Eq. 7.20 is not yet ready for the BOA that will be used to separate the PDE into two ODE systems. The minimum amount of butchering that must be done (if the coordinates ρ and s are not to be mixed to form new coordinates, a step which would complicate the BOA process) is the dropping of the $\partial/\partial s$ term. To determine when this is permissible, consider the ratio of this term to the $\partial^2/\partial s^2$ term. Since we are solving for WG modes, the s-dependence of Ψ (for a ccw traveling wave) is $f(s, \rho)\exp(inks)$ where f varies slowly with s (e.i. $|\partial f/\partial s|/|f| \ll |nk|$). Thus the

magnitude of the ratio of the two terms is approximately

$$\left| \frac{\rho \kappa'}{nk(1 - \rho\kappa)} \right| = \left| \frac{\rho}{R} \frac{1}{1 - \rho\kappa} \frac{R'}{nkR} \right| \tag{7.21}$$

Since Ψ is expected to be exponentially small away from the boundary, we expect

$$\frac{\rho}{R} = \rho\kappa \ll 1. \tag{7.22}$$

We also expect

$$\frac{R'}{nkR} \ll 1 \tag{7.23}$$

because, for a moderately deformed[4] oval resonator, $R' = O(1)$. Thus the ratio (7.21) is made smaller by two factors and this "pre-BOA" approximation is reasonably justified. The Helmholtz equation has now become

$$\left[\frac{\partial^2}{\partial\rho^2} + \frac{-\kappa(s)}{1 - \rho\kappa(s)} \frac{\partial}{\partial\rho} + \frac{1}{(1 - \rho\kappa(s))^2} \frac{\partial^2}{\partial s^2} + n^2 k^2 \right] \Psi = 0. \tag{7.24}$$

We now assume that the wavefunction Ψ can be written as $\Phi(\rho|s)\psi(s)$, where Φ varies rapidly with ρ and slowly with s. The idea is that Φ actually depends on quantities such as $R(s)$ that vary slowly with s, but it is not greatly affected by quantities that vary quickly in s such as the phase. We now place this form of the wave function into our PDE *and neglect the action of $\partial^2/\partial s^2$ on Φ*. This approximation of dropping a derivative with respect to a slow or "massive" degree of freedom in order to separate variables is the defining feature of the generalized Born-Oppenheimer approximation. We now have

$$\psi(s) \left[(1 - \rho\kappa(s))^2 \frac{\partial^2}{\partial\rho^2} - \kappa(s)(1 - \rho\kappa(s)) \frac{\partial}{\partial\rho} + (1 - \rho\kappa(s))^2 n^2 k^2 \right] \Phi(\rho|s)$$
$$= -\Phi(\rho|s) \frac{\partial^2}{\partial s^2} \psi(s) \tag{7.25}$$

Dividing by Ψ yields the BOA separation of variables, in which the separation "con-

---

[4]The amount of geometric deformation will be limited by the subsequent approximation, but here we assume that R' can be order one.

stant", U^2, is a slowly varying function of s:

$$\frac{1}{\Phi}\left[(1-\rho\kappa(s))^2\frac{\mathrm{d}^2}{\mathrm{d}\rho^2} - \kappa(s)(1-\rho\kappa(s))\frac{\mathrm{d}}{\mathrm{d}\rho} + (1-\rho\kappa(s))^2n^2k^2\right]\Phi$$
$$= -\frac{1}{\psi}\frac{\mathrm{d}^2}{\mathrm{d}s^2}\psi$$
$$= U^2(s). \tag{7.26}$$

The ψ equation resulting from the BOA is

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}\psi + U^2(s)\psi = 0. \tag{7.27}$$

This is exactly the relation that we had previously guessed, Eq. (7.1).

7.2.1 The Bessel Form of $\Phi(\rho|s)$

Let's look at the ρ equation (in which s comes in as a parameter via $\kappa(s)$):

$$\left[(1-\rho\kappa(s))^2\frac{\mathrm{d}^2}{\mathrm{d}\rho^2} - \kappa(s)(1-\rho\kappa(s))\frac{\mathrm{d}}{\mathrm{d}\rho} + \left((1-\rho\kappa(s))^2n^2k^2 - U^2(s)\right)\right]\Phi(\rho|s) = 0. \tag{7.28}$$

If we make the substitution

$$y = nk\frac{1-\rho\kappa}{\kappa} = nk(R(s) - \rho), \tag{7.29}$$

then the equation becomes

$$\left[y^2\frac{\mathrm{d}^2}{\mathrm{d}y^2} + y\frac{\mathrm{d}}{\mathrm{d}y} + \left(y^2 - U^2(s)R^2(s)\right)\right]\Phi = 0. \tag{7.30}$$

This is Bessel's equation of (generally complex) order $\nu(s) = RU$. The general solution inside the resonator is

$$\Phi_1(\rho|s) = J_{\nu(s)}(nk(R(s) - \rho)) + C_Y Y_{\nu(s)}(nk(R(s) - \rho)), \tag{7.31}$$

where we have divided out the arbitrary overall amplitude, which will be included in ψ. For a dielectric resonator, the general solution outside of G is

$$\Phi_2(\rho|s) = C_1 H^{(1)}_{\nu(s)}(k(R(s)-\rho)) + C_2 H^{(2)}_{\nu(s)}(k(R(s)-\rho)). \tag{7.32}$$

To examine the validity of the BOA, we can compare terms that have been dropped with terms that have not been dropped from Eq. (7.25), using our above solutions for $\Phi(\rho|s)$. For simplicity we will only consider $\Phi_1(\rho|s)$. Since $Y_\nu(z)$ diverges unphysically at $z = 0$, we must have $C_Y = 0$. The dropped quantity is $\psi(s)\partial^2/\partial s^2 J_{R(s)U(s)}(nkR(s) - nk\rho)$, and performing the derivative results in a number of terms, one of which is

$$\psi(s)[nkR'(s)]^2 J''_{R(s)U(s)}(nkR(s) - nk\rho), \tag{7.33}$$

where the prime indicates the derivative with respect to the entire argument. One of the terms which is kept (the first term of Eq. (7.24) is

$$\psi(s)\frac{\partial^2}{\partial\rho^2} J_{R(s)U(s)}(nkR(s) - nk\rho) = \psi(s)(nk)^2 J''_{R(s)U(s)}(nkR(s) - nk\rho). \tag{7.34}$$

The ratio of the dropped term to the saved term yields the validity condition

$$[R'(s)]^2 \ll 1. \tag{7.35}$$

This is somewhat surprising because it is independent of k. Thus the BOA approximation does not become exact in the limit $k \to \infty$; it is not (solely) a semiclassical approximation.

Some of the other terms which are dropped are more difficult to estimate. We know that $U(s) = O(nk)$, and, if it is further assumed that $U'(s)$ is of the same order as nkR'/R, then the dropped terms are either of the order $O(n^2k^2(R')^2)O(J)O(\psi)$ (such as the one we examined) or of the order $O(nkR'')O(J)O(\psi)$, where $O(J)$ is the order of the Bessel function as well as its first and second derivatives with respect to argument and order. The assumption on the magnitude of U' will be justified in Secs. 7.4.3.

Looking at Eq. (7.24) and keeping in mind that the s derivatives will only act on ψ, we see that the first, third and fourth terms of Eq. (7.24), are of order $O(n^2k^2)O(J)O(\psi)$. Comparing the order of these kept terms with the order of the

dropped terms, we can say that, in order for the BOA to work at all, we need condition (7.35), as well as

$$R''(s)R(s) \ll n|k|R(s), \tag{7.36}$$

Thus the BOA has validity requirements that are both geometric and semiclassical (large kR).

The second kept term (the $\partial/\partial\rho$ term) of Eq. (7.24) is smaller than the other kept terms, and is of order $O(nk/R)O(J)O(\psi)$. In order for the second term to have a significant contribution to accuracy when the BOA is performed, it must be significantly larger than the terms that are dropped. This condition is

$$[R'(s)]^2 n|k|R(s) \ll 1,$$
$$R''(s)R(s) \ll 1. \tag{7.37}$$

When these latter inequalities are *not* met, but the conditions (7.36, 7.35) are met, we have the situation in which the second term might as well be dropped. Even when the inequalities are met, we may wish to drop this term because it is $O(nkR)$ smaller than the other kept terms. This brings us to the Airy form of $\Phi(\rho|s)$.

7.2.2 The Airy Form of $\Phi(\rho|s)$

There are additional approximations made in Refs. [89, 90] that involve dropping parts of Eq. 7.24 and result in the solution for Ψ being expressed in terms of the Airy functions Ai(z) and Bi(z). These functions (of a complex argument in general) are less exotic than Bessel functions of complex *order* and complex argument. Since $\rho/R = \rho\kappa$ is small we expand the $\partial^2/\partial s^2$ term so that it becomes $(1+2\rho\kappa(s))\partial^2/\partial s^2$. In addition, we drop the $\partial/\partial\rho$ term from the differential equation as discussed previously. After then making the BOA separation of variables, the ρ ODE is

$$\frac{d^2}{d\rho^2}\Phi + \left(n^2k^2 - U^2(s) - 2\kappa(s)U^2(s)\rho\right)\Phi = 0 \tag{7.38}$$

Straightforward variable replacements, designed to simplify the coefficient of Φ, lead to the common Airy equation

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2}\Phi - z\Phi = 0, \tag{7.39}$$

where

$$z = (2\kappa U^2)^{1/3}\rho + (2\kappa U^2)^{-2/3}(U^2 - n^2k^2) \tag{7.40}$$

Care must be take when selecting which branch of the cube roots in the above equation should be used (more on this in later sections). The general solution to the Airy equation is

$$\Phi = \mathrm{Ai}(z) + C_B\mathrm{Bi}(z) \tag{7.41}$$

where the arbitrary overall amplitude has been removed here and will be included in ψ. The s ODE when using the Airy form of the approximation is still (7.1).

Unfortunately, as we shall see in Sec. 7.3.4, there are difficulties in using the Airy form with the dielectric resonator.

7.2.3 Boundary Conditions

Now we are ready to set the ODE boundary conditions, which involve additional approximations. The actual boundary conditions for a closed (reflective) resonator are

$$\Phi(0|s)\psi(s) = 0 \quad \forall s$$
$$\psi(s)\Phi(\rho \to R(s)|s) \to 0 \quad \forall s$$
$$\Phi(\rho|\Gamma)\psi(\Gamma) = \Phi(\rho|0)\psi(0) \quad \forall \rho$$
$$\left.\frac{\partial}{\partial s}(\Phi(\rho|s)\psi(s))\right|_{s=\Gamma} = \left.\frac{\partial}{\partial s}(\Phi(\rho|s)\psi(s))\right|_{s=0} \quad \forall \rho, \tag{7.42}$$

Table 7.1: $\psi(s)$ boundary conditions for parity eigenmodes.

parity	BCs due to parity		BC used to set overall amplitude
$(+,+)$	$\psi'(\Gamma/4) = 0$	$\psi'(0) = 0$	$\psi(0) = $ const.
$(+,-)$	$\psi'(\Gamma/4) = 0$	$\psi(0) = 0$	$\psi'(0) = $ const.
$(-,+)$	$\psi(\Gamma/4) = 0$	$\psi'(0) = 0$	$\psi(0) = $ const.
$(-,-)$	$\psi(\Gamma/4) = 0$	$\psi(0) = 0$	$\psi'(0) = $ const.

where Γ is the perimeter of δG. For an dielectric resonator, there are two wavefunctions labeled 1 (inside) and 2 (outside) and the conditions are:

$$\Phi_1(0|s)\psi_1(s) = \Phi_2(0|s)\psi_2(s) \quad \forall s$$

$$\psi_1(s)\frac{\partial}{\partial \rho}\Phi_1(\rho|s)\bigg|_{\rho=0} = \psi_2(s)\frac{\partial}{\partial \rho}\Phi_2(\rho|s)\bigg|_{\rho=0} \quad \forall s$$

$$\psi_1(s)\Phi_1(\rho \to R(s)|s) \to 0 \quad \forall s$$

$\psi_2(s)\Phi_2(\rho \to -\infty|s)$ should not be an incoming wave $\forall s$

$$\Phi_i(\rho|\Gamma)\psi_i(\Gamma) = \Phi_i(\rho|0)\psi_i(0) \quad \forall \rho$$

$$\frac{\partial}{\partial s}(\Phi_i(\rho|s)\psi_i(s))\bigg|_{s=\Gamma} = \frac{\partial}{\partial s}(\Phi_i(\rho|s)\psi_i(s))\bigg|_{s=0} \quad \forall \rho. \tag{7.43}$$

At this point, we will assume that we can make the natural factorization (separation) of these boundary conditions so that they may be used in the ODEs for ψ and Φ. The $\psi(s)$ BCs are then

$$\psi(\Gamma) = \psi(0), \tag{7.44}$$

$$\psi'(\Gamma) = \psi'(0). \tag{7.45}$$

(For the dielectric resonator, we take $\psi_1 = \psi_2 \equiv \psi$.)

The $\psi(s)$ boundary conditions can be taken at $\Gamma/2$ or $\Gamma/4$ if we make use of symmetry. The true modes of the resonator should be eigenfunctions of x-parity and y-parity, as explained in the group theory discussion in Ref. [34]. The four possible parity states that a mode can have are $(+,+)$, $(+,-)$, $(-,+)$, $(-,-)$, where $(+,-)$ denotes a mode having x-parity 1 (even in reflections about the y axis) and y-parity -1 (odd in reflections about the x axis). Therefore, numerical techniques for solving the ψ ODE may be able to use boundary conditions on the quarter resonator as given in Table 7.1.

The Φ BCs for the closed resonator are

$$\Phi(0|s) = 0 \quad \forall s \tag{7.46}$$

$$\Phi(\rho \to R(s)|s) \to 0 \quad \forall s, \tag{7.47}$$

and the Φ BCs for the dielectric resonator are

$$\Phi_1(0|s) = \Phi_2(0|s) \quad \forall s \tag{7.48}$$

$$\Phi_1'(0|s) = \Phi_2'(0|s) \quad \forall s \tag{7.49}$$

$$\Phi_1(\rho \to R(s)|s) \to 0 \quad \forall s \tag{7.50}$$

$$\Phi_2(\rho \to -\infty|s) \quad \text{should not be an incoming wave} \quad \forall s. \tag{7.51}$$

7.3 Exact Calculation of $U(s)$

Here we review the equations that must be solved in order to exactly calculate $U(s)$.

7.3.1 Reflective Resonator, Bessel Form

The BC (7.47) allows us to set $C_Y = 0$ in (7.31). The BC (7.46) now yields

$$J_{R(s)U(s)}\big(nkR(s)\big) = 0 \tag{7.52}$$

The innermost process in determining the solution involves numerically solving this equation for U given s and k. Let's call the routine that performs this GET_U(s, k). Since k and R are real for the closed resonator, U will always be real (this is not the case for an open resonator, since k is complex). The next step is to solve (7.1) for $\psi(s)$ via the WKB approximation or by numerical ODE methods. Either of these solution processes will call GET_U(s, k) many times.

7.3.2 Reflective Resonator, Airy Form

The BC (7.47) tells us that as ρ increases, Φ must rapidly decay before reaching R. Since the z in (7.41) increases with ρ, and Ai(z) super-exponentially decays beyond a certain point ($z = 0$), we see that the Ai function has the correct behavior. (The wavefunction decays at the internal caustic of the immediate ray motion, and indeed, and Ai function is exactly what is used to describe behavior at a simple fold caustic

Figure 7.1: The Airy functions.
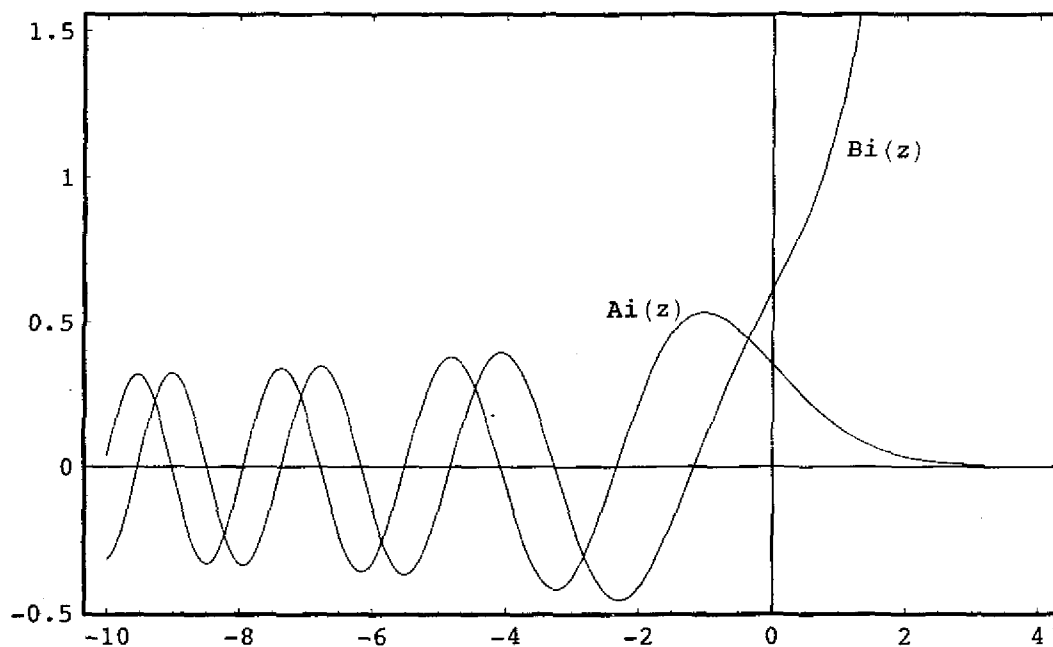
(turning-point) in the standard semiclassical methods). The Bi function grows super-exponentially for $z > 0$ and therefore $C_B = 0$ in the solution (7.41). The Ai and Bi functions are shown in Fig. 7.1. Now, using the BC (7.46) we have

$$\mathrm{Ai}\big((2\kappa(s)U^2(s))^{-2/3}(U(s) - k^2)\big) = 0. \tag{7.53}$$

The zeros of the Ai function, z_n, are tabulated, and are all real and negative. Thus we should be able to take $U(s)$ to be real (k is real for a closed resonator). Since

$$(2\kappa(s)U^2(s))^{-2/3}(U^2(s) - k^2) = z_n < 0 \tag{7.54}$$

Only the last of the three possibilities—($U^2 > 0$, $U^2 - k^2 > 0$), ($U^2 < 0$, $U^2 - k^2 < 0$), and ($U^2 > 0$, $U^2 - k^2 < 0$)— will work. Now we define

$$u \equiv U^{2/3} \tag{7.55}$$

and know that $u > 0$. We obtain the cubic equation

$$u^3 + \left(-z_n(2\kappa)^{2/3}\right)u^2 - k^2 = 0 \tag{7.56}$$

Playing graphically with the function $x^3 + px^2 - q$ with $p > 0, q > 0$ reveals there is always exactly one positive real zero. The three solutions of $x^3 + px^2 - q = 0$ are

$$x = \frac{1}{3}\left[-p + \frac{2^{1/3}p^2}{\zeta}\left\{\begin{array}{c}1\\ \exp(-i\pi2/3)\\ \exp(i\pi2/3)\end{array}\right\} + 2^{-1/3}\zeta\left\{\begin{array}{c}1\\ \exp(i\pi2/3)\\ \exp(-i\pi2/3)\end{array}\right\}\right] \tag{7.57}$$

where

$$\zeta = \left((27q - 2p^3) + \sqrt{27q(27q - 4p^3)}\right)^{1/3} \tag{7.58}$$

Since $p^3/q = 4|z_n|^2/(kR)^2$ and we take kR to be sufficiently large, we see that the square root and the cube root in the equation above are always acting on positive real numbers. Thus $\zeta > 0$ and we know that the first solution in (7.57) is real valued and therefore must be the correct one (which must be positive):

$$u = \frac{1}{3}\left(-p + \frac{2^{1/3}p^2}{\zeta} + 2^{-1/3}\zeta\right) > 0. \tag{7.59}$$

Thus for the closed resonator, in the Airy form of the approximation, the GET_U(s, k) routine is the calculation of the analytic algebraic expression for $U = u^{3/2}$.

We can use the Airy form of $\Phi(\rho|s)$ to get an estimate of the spatial thickness of the whispering gallery mode. The nominal position of the inner caustic occurs where the argument of Ai is zero, as for arguments larger than this, the function rapidly decays. Including an offset of $+0.66$ gives the location where the field value of the inner caustic has decayed to $1/e$ of its maximum value. Including an offset of -1.02 gives the location where the field value of the innermost radial antinode has a maximum value. Looking again at Eq. (7.40), and noting that the second term on the right hand side is z_n, we see that the caustic position occurs at

$$\rho_{\rm c} \approx (2\kappa U^2)^{-1/3}(z_{\rm off} - z_n), \tag{7.60}$$

or

$$\frac{\rho_c(s)}{R(s)} \approx \frac{0.79(z_{\text{off}} - z_n)}{\nu^{2/3}(s)},$$
(7.61)

where z_{off} is 0, 0.66, or -1.02 depending on the desired definition of the caustic. The values of z_1 through z_7 are -2.34, -4.09, -5.52, -6.79, -7.94, -9.02, -10.04.

7.3.3 Dielectric Resonator, Bessel Form

For the dielectric resonator we will denote the refractive index inside the resonator to be n and set the index outside to be 1. The BC (7.50) forces us to drop the Neumann function part of Φ_1, since $Y_\nu(z)$ diverges as z goes to zero. Thus we have

$$\Phi_1(\rho|s) = J_{RU}(nk(R - \rho)).$$
(7.62)

The BC (7.51) requires that Φ_2 have no overlap with $H^{(2)}$, the incoming type of Hankel function. Thus Φ_2 must be an outgoing Hankel function $H^{(1)} = J + iY$:

$$\Phi_2(\rho|s) = C_1 H_{RU}^{(1)}[k(R - \rho)].$$
(7.63)

Here C_1 is an arbitrary constant that must be determined.

The BCs (7.48, 7.49) yield

$$J_{R(s)U(s)}(nkR(s)) = C_1 H_{R(s)U(s)}^{(1)}(kR(s)),$$
(7.64)

$$nJ'_{R(s)U(s)}(nkR(s)) = C_1 H_{R(s)U(s)}^{(1)\prime}(kR(s)).$$
(7.65)

The unknown coefficient C_1 can be eliminated to yield

$$J_{R(s)U(s)}(nkR(s))H_{R(s)U(s)}^{(1)\prime}(kR(s)) = nJ'_{R(s)U(s)}(nkR(s))H_{R(s)U(s)}^{(1)}(kR(s)),$$
(7.66)

which simplifies to Eq. (7.2) upon use of the Bessel function recursion relations. Eq. (7.2) must be solved numerically for complex U (this is the task of the GET_U(s, k) routine). Since the resonator is open, the input parameter k will have a negative imaginary part.

7.3.4 No Solution: Dielectric Resonator, Airy Form

Although one would think that the Airy form solution of the dielectric resonator would be simpler than the Bessel form, it seems there is no good way to treat BC (7.51) which stipulates that there be no incoming wave. A possible alternative to (7.51) is to require that Φ_2 decay outside the resonator. In certain parameter regimes, including a Bi function in Φ_2 can cause Φ_2 to decay to a small amplitude and then oscillate with small oscillations out to infinity. This is a somewhat physically intuitive result, but this approach seems to give no constraint on the ratio of the coefficient of Bi to the coefficient of Ai in the expression for Φ_2. In the one implementation that has been tried, we have set $\Phi_2 = \text{Bi}(z(n))$, completely dropping the Ai part. As for Φ_1, the BC (7.47) requires that $\Phi_1 = \text{Ai}(z(n = 0))$. The agreement of this approach with the Bessel form has yet to be numerically analyzed.

The problem with the boundary condition may come from the fact that the approximations that led to the Airy form restrict us to small ρ. Since this is true outside as well as inside the resonator, it seems problematic to use any boundary condition that involves $\rho \to -\infty$. This objection can be raised for the Bessel form of Φ as well, although there is no obvious roadblock to using the Bessel form as there is for the Airy form. Perhaps this is related to the fact that the approximations made in deriving the Bessel function form do not rely so heavily on $|\rho| \ll R$.

7.4 Approximating $\psi(s)$

For the rest of the chapter we return specifically to the case of a dielectric resonator, where the Bessel form of GET_U must be used.

7.4.1 The WKB approximation

The standard presentation of the WKB approximation involves the 1D time-independent Schrödinger equation,

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi + 2m(E - V(x))\hbar^2\psi = 0, \qquad (7.67)$$

and expands the natural logarithm of $\psi(x)$ in powers of $\hbar$. While there is no $\hbar$ in Eq. (7.1), one can nevertheless simply use the WKB result by comparing this equation with the Schrödinger equation and setting $\hbar = 1$ at the end. If this seems

unsatisfactory, one can use the iterative perturbation approach, used in [92], which omits the use of $\hbar$.[5] We present the iterative perturbation theory here.

We substitute for ψ the function ϕ, defined by

$$\psi(s) \equiv e^{i\phi(s)}. \tag{7.68}$$

Inserting into Eq. (7.1) yields

$$i\phi'' - (\phi')^2 + U^2 = 0, \tag{7.69}$$

which, with the substitution $y(s) \equiv \phi'(s)$, can be written as

$$y^2 = U^2 + iy'. \tag{7.70}$$

We assume that

$$|y'| \ll U^2, \tag{7.71}$$

so that from (7.70) we have the first approximation for y:

$$y \approx \pm U. \tag{7.72}$$

Taking the derivative yields

$$y' \approx \pm U'. \tag{7.73}$$

Substituting (7.73) this into the RHS of (7.70) allows us to calculate the next (WKB) approximation for y:

$$\begin{aligned} y &\approx \pm\sqrt{U^2 \pm iU'} \\ &\approx \pm U\sqrt{1 \pm i\frac{U'}{U^2}} \\ &\approx \pm U + i\frac{U'}{2U}. \end{aligned} \tag{7.74}$$

(The factoring of the square root is valid because $\operatorname{Re} U > 0$.) We see that the validity

---

[5]A more formal approach is to use multiple-scale analysis [93], which does use a small parameter, and essentially includes WKB as a simple case.

condition for the WKB approximation is

$$\frac{U'}{U^2} \ll 1. \tag{7.75}$$

Integrating y from (7.74) to obtain ϕ yields

$$\phi \approx \pm \int^s U(s')\,\mathrm{d}s' + \frac{i}{2}\ln|U| + \text{const.} \tag{7.76}$$

Substituting ϕ into ψ gives the WKB result:

$$\psi(s) \approx \frac{A}{\sqrt{|U(s)|}}\exp\left[i\int^s U(s')\,\mathrm{d}s'\right] + \frac{B}{\sqrt{|U(s)|}}\exp\left[-i\int^s U(s')\,\mathrm{d}s'\right], \tag{7.77}$$

where A and B are constants.

7.4.2 The Resonator Picture, the Waveguide Picture, and Effective Paths

As hinted at in the beginning of this chapter, there are at least two pictures of the BOA-separated WG wave. In first picture, which we can call the resonator picture and will denote by a superscript (1), complex $k^{(1)}$ is adjusted so that $\psi^{(1)}$ reconstructs itself exactly after one trip around the perimeter of the resonator. In this picture $U^{(1)}(s)$ is generally complex, but its imaginary part oscillates between being positive in the regions of small radius of curvature (where tunneling transmittance is the greatest) and being negative in the regions of large radius of curvature, where the tunneling transmittance is smallest. For the resonant value $k = k^{(1)}$, the average of $\operatorname{Im} U^{(1)}$ is zero:

$$\langle \operatorname{Im} U^{(1)}(k^{(1)}) \rangle = \frac{1}{\Gamma}\int_0^\Gamma \operatorname{Im} U^{(1)}(s, k^{(1)})\,\mathrm{d}s = 0. \tag{7.78}$$

For a circular resonator where R is constant, we find indeed that the solutions have real $U^{(1)}$, that is, the angular momentum is real and the wavenumber is complex. (When the periodic boundary condition on $\psi^{(1)}(s)$ is enforced, the angular momentum $RU^{(1)}$ is, of course, an integer.)

The other picture can be called the waveguide picture. In this picture we imagine a wave traveling along a curved dielectric interface. The wavenumber $k^{(2)}$ is taken

to be real and equal to ω/c. It is natural to set $k^{(2)} = \omega/c$, a real number. As the wave propagates, it decreases in magnitude due to a strictly positive imaginary part of $U^{(2)}$. (The resonance condition enforces $\operatorname{Im} U > 0$ when k is real). One can show that, when the WKB form of ψ is used, a traveling wave undergoes no reflection, only loss due to tunneling. Thus, if a traveling wave propagates from $s = s_1$ to $s = s_2$, where $R(s_2) = R(s_1)$ but R can vary on (s_1, s_2), the power transmittance of the wave across the interval in the WKB approximation is simply

$$T_{12} = \exp\left(-2\operatorname{Im}\int_{s_1}^{s_2} U^{(2)}(s, k)\,\mathrm{d}s\right), \tag{7.79}$$

as stated in Eq. (7.5).

Now imagine closing the waveguide so that in fact it is an oval resonator. In the waveguide picture the wave decreases in amplitude by $\exp[-\operatorname{Im}\int_0^\Gamma U^{(2)}(s, \omega/c)\,\mathrm{d}s]$ in one round trip. The waveguide picture loss rate is $\langle\operatorname{Im} U^{(2)}(\omega/c)\rangle$. It is interesting to compare (by ratio) this average loss rate with the resonator picture loss rate (or, more accurately, gain rate) for propagation within the resonator medium, $-n\operatorname{Im} k^{(1)}$. Another quantity of interest is the ratio of $\langle\operatorname{Re} U^{(2)}(\omega/c)\rangle$, the average of the real part of $U^{(2)}(s, \omega/c)$, to $n\operatorname{Re} k^{(1)}$. We will look at some values of these ratios for the circle, where the averages over s are not needed.

In the classical ray picture, a wave travels along a trajectory within the resonator. As it travels, we can presume that its phase advances at the rate $n\operatorname{Re} k^{(1)}$ and its amplitude grows with rate $-n\operatorname{Im} k^{(1)}$. The perimeter of a regular polygon inscribed in the unit circle is $2\pi \times \cos(\chi)/(\pi/2 - \chi)$, where χ corresponds to the angle of incidence of the polygon trajectory ($\chi = \pi/2 - \pi/n$ for the n-gon). Extending this perimeter formula to non-closed orbits (still in the circle), we can say that the path length of a "round trip" trajectory is less than the circumference $\Gamma = 2\pi$, by a factor of about $\cos(\chi)/(\pi/2-\chi)$. One might postulate then that the ratios described in the previous paragraph are in fact both equal to (or near to) $\cos(\chi)/(\pi/2-\chi)$. The results we obtain next will show that the phase advance needs to be modified (retarded)— this could already be presumed by noting that the beam is better represented by a Gaussian than a plane wave, and that a Gaussian beam has the retarding Gouy phase shift. Instead of retarding the phase advance, we will find a shorter effective path.

Another model one might consider is that the light travels along a smooth effective path (closed curve) which lies inside the billiard and is described by a local radius $r_{\mathrm{eff}}(s)$, which is measured from the local center of curvature. The idea is motivated

by the fact that in the semiclassical solution of the circle [75], the wavefronts of the semiclassical wavefunction are perpendicular to the caustic at the location of the caustic, and thus one may think of a "plane wave" traveling along the curved caustic as an alternative picture to thinking of plane waves traveling along the bouncing classical trajectory.

To quantify the comparison, imagine that we have a resonator picture solution, with complex wavenumber $k^{(1)}$ and real angular momentum $\nu^{(1)}$. For ease of referring to the ratios previously mentioned, we define the notation

$$\frac{r_{\text{eff}}^{(\text{I})}(s)}{R(s)} \equiv \frac{\operatorname{Im} U^{(2)}(s, \omega/c)}{-n \operatorname{Im} k^{(1)}}, \tag{7.80}$$

$$\frac{r_{\text{eff}}^{(\text{R})}(s)}{R(s)} \equiv \frac{\operatorname{Re} U^{(2)}(s, \omega/c)}{n \operatorname{Re} k^{(1)}}. \tag{7.81}$$

The s dependence has been included because we will use these ratios for noncircular billiards as well.

We first examine $r_{\text{eff}}^{(\text{R})}$. For high Q resonators, it can be expected that $\operatorname{Re} U^{(2)} \approx \operatorname{Re} U^{(1)}$. (It is not that important which superscript of U is used in definition (7.81).) $\operatorname{Re} U^{(1)} R$ is $\operatorname{Re} \nu^{(1)}$, the real part of the local angular momentum. This quantity is approximately equal to $n \operatorname{Re} k r_{\text{caustic}}(s)$, where $r_{\text{caustic}}(s)$ is the distance from the local center of curvature, $O(s)$, to the intersection of the classical caustic with the radial line going from $O(s)$ to the billiard point specified by s. This is true in as much as the classical caustic is perpendicular to the radial line (or in as much as it is locally parallel to the billiard boundary). Using definition (7.81) now yields

$$r_{\text{eff}}^{(\text{R})}(s) \approx r_{\text{caustic}}(s). \tag{7.82}$$

Thus $r_{\text{eff}}^{(\text{R})}$ is, very loosely speaking, the minimum radius of the classical trajectory, not the "average" radius of $R \cos \chi / (\pi/2 - \chi)$. We can say that it is the classical caustic, not the classical trajectory, that is the effective path for phase advance, e.i. the phase advance along the billiard surface due to the "local wavenumber" $\operatorname{Re} U$ is approximately equal to the phase advance due to the global wavenumber $n \operatorname{Re} k$ along the caustic.

We now examine $r_{\text{eff}}^{(\text{I})}$. In doing so we will explicitly define the constraints yielding the solution in both pictures. We define complex F as being the quantity which is

zero when the resonance condition is met:

$$F(x, \nu) = J_\nu(nx) H_{\nu-1}^{(1)}(x) - n J_{\nu-1}(nx) H_\nu^{(1)}(x), \tag{7.83}$$

where it is understood that $x \equiv kR(s)$. For future reference we note that

$$\frac{\partial}{\partial x} F(x, \nu) = (n^2 - 1) J_\nu(nx) H_\nu^{(1)}(x) - \frac{F}{x}. \tag{7.84}$$

Our resonator picture solution is determined by

$$F(x^{(1)}, \nu^{(1)}) = 0, \tag{7.85}$$

$$\psi^{(1)}(\Gamma) = \psi^{(1)}(0). \tag{7.86}$$

The waveguide picture solution is defined by

$$F(\operatorname{Re} x^{(1)}, \nu^{(2)}) = 0; \tag{7.87}$$

that is, we have set $k^{(2)} = \operatorname{Re} k^{(1)}$ and solved for $\nu^{(2)}$. The boundary wave $\psi^{(2)}$ is then calculated from $k^{(2)}$ and $\nu^{(2)}$. $\psi^{(2)}$ does *not* meet periodic boundary conditions.

We wish to learn something about $i\operatorname{Im} k^{(1)} = k^{(1)} - k^{(2)}$. We can obtain the latter form by linearly expanding $F(x, \nu)$ about $(x^{(1)}, \nu^{(1)})$ and evaluating the expansion at $(x^{(2)}, \nu^{(2)})$:

$$F(x^{(2)}, \nu^{(2)}) \approx F(x^{(1)}, \nu^{(1)}) + \left(\frac{\partial F(x^{(1)}, \nu^{(1)})}{\partial x} \right)_\nu (x^{(2)} - x^{(1)})$$
$$+ \left(\frac{\partial F(x^{(1)}, \nu^{(1)})}{\partial \nu} \right)_x (\nu^{(2)} - \nu^{(1)}). \tag{7.88}$$

The truncation of the power series at linear order is assumed to be valid because $\operatorname{Im} x^{(1)} \ll \operatorname{Re} x^{(1)}$. Since both $F(x^{(1)}, \nu^{(1)})$ and $F(x^{(2)}, \nu^{(2)})$ are zero, we have

$$\frac{\nu^{(2)} - \nu^{(1)}}{x^{(2)} - x^{(1)}} \approx - \frac{\left(\frac{\partial F(x^{(1)}, \nu^{(1)})}{\partial x} \right)_\nu}{\left(\frac{\partial F(x^{(1)}, \nu^{(1)})}{\partial \nu} \right)_x}. \tag{7.89}$$

Note that the RHS is exactly equal to

$$\left(\frac{\partial \nu(x^{(1)}, F = 0)}{\partial x} \right)_F \equiv \frac{d\nu^{(1)}}{dx^{(1)}}. \tag{7.90}$$

The LHS of (7.89) is equal to

$$\frac{U^{(2)}(s) - U^{(1)}(s)}{-i\operatorname{Im} k^{(1)}},\tag{7.91}$$

which is approaching the desired form of $\operatorname{Im} U / \operatorname{Im} k$.

Noting that the imaginary parts of ν and x are much smaller than their respective real parts, an application of the first order Debye approximation (see Appendix H) to the two partial derivatives in Eq. (7.89) results in the fact that, if $|1.9\nu(\nu/x-1)^{3/2}| \gg 1$, then both of the partial derivatives are nearly (relatively) pure imaginary (because $|Y_\nu| \gg |J_\nu|$). Comparing with (7.91), we see that $U^{(2)} - U^{(1)}$ must have a relatively negligable real part:

$$U^{(2)} - U^{(1)} \approx \operatorname{Im} U^{(2)} - \operatorname{Im} U^{(1)}.\tag{7.92}$$

Finally we note that, in the circle, $\operatorname{Im} U^{(1)} = 0$, while for a general oval $\langle \operatorname{Im} U^{(1)} \rangle = 0$.

Thus we have the desired ratio

$$\left\langle \frac{r_{\text{eff}}^{(\text{I})}}{R} \right\rangle = \frac{-\langle \operatorname{Im} U^{(2)} \rangle}{n \operatorname{Im} k^{(1)}} \approx \frac{1}{n} \left\langle \frac{d\nu^{(1)}}{dx^{(1)}} \right\rangle$$
$$\approx \left(\frac{1}{n} - n \right) \left\langle J_{\nu^{(1)}}(nx^{(1)}) H^{(1)}_{\nu^{(1)}}(x^{(1)}) \left(\frac{\partial F(x^{(1)}, \nu^{(1)})}{\partial \nu} \right)_x^{-1} \right\rangle,\tag{7.93}$$

where we have used Eq. (7.84) in the last step. If desired, a first order Debye approximation for $\partial F / \partial \nu$, in terms of Bessel functions, can be found by piecing together (7.83, H.3, H.4).

In general, we have found that $r_{\text{eff}}^{(\text{I})} \neq r_{\text{eff}}^{(\text{R})}$. The former is proportional to the derivative $(\partial \nu / \partial x)_{F=0}$ while the latter is proportional to the direct ratio $(\nu/x)_{F=0}$, and is shown to be the caustic radius in the limits of small deformation from the circle and large Q.

Direct computation of $r_{\text{eff}}^{(\text{R})}$ and $r_{\text{eff}}^{(\text{I})}$ by the defining ratios (7.80, 7.81) for several different resonances in the circle are shown in Table 7.2. The data shows that, for these parameters, $r_{\text{eff}}^{(\text{I})}/R$ is relatively close to $\cos\chi/(\pi/2 - \chi)$, where χ is chosen to satisfy $r_{\text{eff}}^{(\text{R})}/R = \sin\chi$. This suggests that the trajectory model, in which a beam follows the classical trajectory and experiences constant gain in order to offset the transmission loss at the bounces, has a substantial degree of validity. We state the

kR	n	m	$r_{\text{eff}}^{(\text{R})}/R$	$r_{\text{eff}}^{(\text{I})}/R$	$\cos\chi/(\pi/2-\chi)$
$53.4372 - 0.10603i$	1.5	55	0.686	0.882	0.893
$54.3593 - 0.010970i$	1.5	60	0.736	0.898	0.910
$54.2282 - 2.0622 \times 10^{-4}i$	1.5	64	0.787	0.919	0.928
$55.0937 - 0.01467i$	3.0	57	0.345	0.764	0.770
$54.6970 - 1.4879 \times 10^{-9}i$	3.0	72	0.439	0.803	0.805

Table 7.2: Effective radii for several circle resonances.

result for the circle

$$\text{Im}U^{(2)}\Gamma = -n\text{Im}k^{(1)} \times (\text{trajectory round trip length}) \qquad (7.94)$$

This loss-gain balance is quite similar to that used the sequential tunneling model, which was shown to do quite well in Fig. 5.2. For noncircular resonators, the left hand side must be averaged to get rid of its dependence on s (the RHS does not depend on s). This can be expressed as

$$\int_0^{\Gamma} \text{Im}U^{(2)} \, ds = -n\text{Im}k^{(1)} \times (\text{trajectory round trip length}). \qquad (7.95)$$

A numerical calculation supporting this equality will be given in Sec. 7.5.3.

The results of this subsection suggest the following: For the phase propagation, imagine a "plane wave" of wavenumber $\text{Re}\,k^{(1)}$ traveling along the caustic; the phase advance in one round trip is equal to the phase advance $\int \text{Re}U \, ds$ along the boundary. In other words, the effective path for the phase propagation is simply the caustic itself.

On the other hand, for amplitude evolution, the correct effective path to use appears to be the true classical trajectory (in which the caustic is inscribed). This statement is shown by Eq. (7.95), in which the "trajectory round trip length" refers to the length of the round trip of the classical trajectory.

7.4.3 Is WKB as Good as BOA?

It is of interest to determine whether performing the WKB approximation to solve the ψ ODE (7.1) involves only the same level of approximations that went into the BOA. If this is the case, there is no significant benefit in solving (7.1) by numerical techniques such as Runga-Kutta methods; the WKB approximation can always be used as an extension of the BOA process.

As given in (7.75), the validity of the WKB approximation depends on the relation

$$\frac{1}{U^2}\left(\frac{\partial U}{\partial s}\right)_{k,F=0} \ll 1. \tag{7.96}$$

The use of partial derivatives and the omission of averaging brackets transforms this condition into

$$\frac{1}{|U(s)|}\frac{|r_{\text{eff}}^{(I)}(s) - r_{\text{eff}}^{(R)}(s)|}{r_{\text{eff}}^{(R)}(s)}\frac{R'(s)}{R(s)} \ll 1. \tag{7.97}$$

Note that since the middle factor is less than one and $U = O(nk)$, we see that U' is of order $O(nkR'/R)$ or less, as was assumed near the end of Sec. 7.2.1.

This relation can be compared to the pre-BOA condition,

$$\frac{1}{n|k|}\frac{\rho}{R(s)}\frac{R'(s)}{R(s)} \ll 1, \tag{7.98}$$

discussed in the beginning of Sec. 7.2. The maximum ρ is roughly equal to $R - r_{\text{eff}}^{(R)}$. We now make this substitution in (7.98), and compare to the WKB validity condition (7.97). We conclude that if

$$\frac{|r_{\text{eff}}^{(I)}(s) - r_{\text{eff}}^{(R)}(s)|}{r_{\text{eff}}^{(R)}(s)} < \frac{R - r_{\text{eff}}^{(R)}}{R}(s), \tag{7.99}$$

the WKB approximation might as well be made whenever the BOA procedure is used, because its validity condition is automatically met when the BOA procedure is valid. The trials shown in Table 7.2 having $n = 1.5$ satisfy inequality (7.99), while those having $n = 3$ do not. Of course these trials are done for a circle and do not handle the variation in r_{eff} with s for a non-circular billiard (this comes from the variation in U which itself comes from the variation in R). We cannot state exactly the parameter regime for which the validity of BOA implies the validity of WKB. However, it seems unlikely that a large, realistic parameter regime exists in which the BOA procedure is demonstratively a very good approximation yet the addition of the WKB approximation renders the result a very poor approximation.

7.5 Application to the Dielectric Ellipse

In this section we numerically calculate the BO-WKB solution and the exact solutions for the dielectric ellipse.

7.5.1 Transformation to Practical Coordinates

The coordinate s, like Elmer Fudd in the 2003 Looney Tunes movie, just turns out to be secretly evil. Although $R(s)$ is the natural parameterization of any curve, the form of $R(s)$ is not simple for a number of commonly used billiard shapes such as the ellipse and the quadrupole. For the ellipse, $R(\theta_f)$ is known in closed form, where the "f" subscript means that θ_f is the angular coordinate with respect an the origin located at one focus of the ellipse. Figure 7.2 depicts different coordinates we will be using in this section. While $s(\theta_f)$ involves elliptical integrals, $ds/d\theta_f$ is simple and turns out to be all that we need to turn the ODE for $\psi(s)$ (7.1) into an ODE for $\psi(\theta_f)$. Whether or not the WKB approximation is used, s is no longer needed, anywhere, and the entire problem is solved with the θ_f and ρ variables. It is important to remember that θ_f is the angular coordinate of a point on the boundary and not the angular coordinate of a general point on the plane; θ_f is a function of s only, not ρ. In principle a similar procedure can be done for the quadrupole, in which case one would use the parametrization $r(\theta) = r_0 + r_0\epsilon\cos(2\theta)$ (for which the origin is at the center of the billiard/resonator).

It is well known that the ellipse can be parametrized by

$$r_f(\theta_f) = \frac{a(1-e^2)}{1+e\cos\theta_f}, \tag{7.100}$$

where r_f is the distance from the focus we have chosen to a point on the ellipse. (It may also be parametrized with the origin at the center of the ellipse by $r(\theta) = a[(1-e^2)/(1-e^2\cos^2\theta)]^{1/2}$. The (r,θ) coordinates will be used later. Some properties in this coordinate system are given in Appendix I.) In [75], we find that the radius of curvature at a point on the ellipse is given by

$$R(r_f) = \frac{\left(r_f(2a-r_f)\right)^{3/2}}{ab}. \tag{7.101}$$

We use the standard notation where a is the semimajor axis, b is the semiminor axis, $c = (a^2-b^2)^{1/2}$ is the distance from the center of the ellipse to a focus, and $e = c/a$

Figure 7.2: Some coordinates specifying the location of a point on an ellipse. The center and foci of the ellipse are marked with dots.
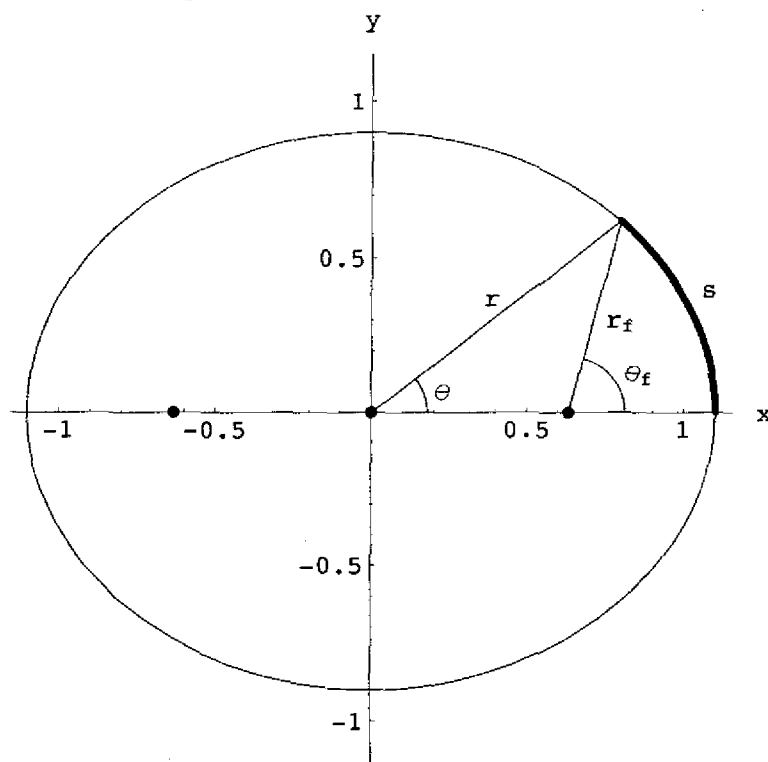
is the eccentricity. Combining the above two equation yields $R(\theta_f)$ and $\kappa(\theta_f)$, which can be used in place of $R(s)$ and $\kappa(s)$ in previous expressions such as the resonance condition (7.2). This results in the GET_U routine being a numerical function of k and θ_f.

At this point we can write the (ccw traveling) WKB solution as

$$\psi(\theta_f) = U(\theta_f)^{-1/2} \exp\left(i \int^{\theta_f} U(\theta_f') \frac{ds(\theta_f')}{d\theta_f'} \, d\theta_f'\right). \tag{7.102}$$

The s derivative in the integrand will be calculated shortly.

If one chooses not to make the WKB approximation, one can proceed as follows. It is relatively simple to derive the general transformation of a second (non-partial)

derivative from a variable x to a variable y when $x(y)$ is known:

$$\frac{\mathrm{d}}{\mathrm{d}x} = \left(\frac{\mathrm{d}x}{\mathrm{d}y}\right)^{-1}\frac{\mathrm{d}}{\mathrm{d}y},$$
$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} = \left(\frac{\mathrm{d}x}{\mathrm{d}y}\right)^{-2}\frac{\mathrm{d}^2}{\mathrm{d}y^2} - \left(\frac{\mathrm{d}x}{\mathrm{d}y}\right)^{-3}\left(\frac{\mathrm{d}^2x}{\mathrm{d}y^2}\right)\frac{\mathrm{d}}{\mathrm{d}y}. \tag{7.103}$$

From this we see that the ψ ODE (7.1) becomes

$$\left[\frac{\mathrm{d}^2}{\mathrm{d}\theta_f^2} - \left(\frac{\mathrm{d}s}{\mathrm{d}\theta_f}\right)^{-1}\left(\frac{\mathrm{d}^2s}{\mathrm{d}\theta_f^2}\right)\frac{\mathrm{d}}{\mathrm{d}\theta_f} + \left(\frac{\mathrm{d}s}{\mathrm{d}\theta_f}\right)^2 U(\theta_f)\right]\psi(\theta_f) = 0. \tag{7.104}$$

The next step is to find the ψ ODE in the variable θ_f. We can directly find $\mathrm{d}s/\mathrm{d}\theta_f$ by using the total differential equation $\mathrm{d}s = (\mathrm{d}r_f^2 + r_f^2\mathrm{d}\theta_f^2)^{1/2}$. Thus

$$\frac{\mathrm{d}s}{\mathrm{d}\theta_f} = \sqrt{\left(\frac{\mathrm{d}r_f}{\mathrm{d}\theta_f}\right)^2 + r_f^2}$$
$$= \frac{a(1 - e^2)\sqrt{1 + e^2 + 2e\cos\theta_f}}{(1 + e\cos\theta_f)^2}, \tag{7.105}$$
$$\frac{\mathrm{d}^2s}{\mathrm{d}\theta_f^2} = \frac{ae(1 - e^2)\sin\theta_f}{(1 + e\cos\theta_f)^2}\left(\frac{2\sqrt{1 + e^2 + 2e\cos\theta_f}}{1 + e\cos\theta_f} - \frac{1}{\sqrt{1 + e^2 + 2e\cos\theta_f}}\right). \tag{7.106}$$

Substituting these two quantities into (7.104) yields the ψ ODE for the ellipse. The first derivative can also be substituted into (7.102) to obtain the WKB solution. We note that Eqs. (7.100, 7.101, 7.105) can be used to show that maximum of the geometric quantity $|R'(s)|$, which must be small, per strong inequality (7.35), in order for the BOA to be valid, is given by

$$\sup_{s\in[0,\Gamma)} |R'| = \frac{3}{2}\frac{e^2}{\sqrt{1 - e^2}} = \frac{3}{2}\frac{c^2}{ab} = \frac{3}{2}\left(\frac{a}{b} - \frac{b}{a}\right). \tag{7.107}$$

If this quantity is too large, the BOA cannot be used no matter how small one makes wavelength.

For completeness we give $s(\theta_f)$ here, which involves the tabulated incomplete elliptic integral of the second kind, $E(\phi|m)$.

$$s(\theta_f) = \frac{\Gamma}{4} - aE(\arcsin[x(\theta_f)/a]|e^2), \quad 0 \leq \theta_f \leq \pi - \arccos(e), \tag{7.108}$$

where $x(\theta_f) = ae + r_f(\theta_f)\cos(\theta_f)$ is the x coordinate with respect to an origin located at the *center* of the ellipse, and Γ is the full perimeter of the ellipse given by

$$\Gamma = s(2\pi) = 4aE(\pi/2|e^2) = 4aE(e^2), \tag{7.109}$$

the latter function $E(m)$ being the *complete* elliptic integral of the second kind. Note that $\theta_f = \pi - \arccos(e)$ corresponds to one quarter of the way around the ellipse.

If one avoids the WKB approximation, the ψ ODE is numerically integrated from $\theta_f = 0$ to $\theta_f = \pi - \arccos(e)$ (one quarter of the way around the billiard) starting with the initial conditions at $\theta_f = s = 0$ given in one of the rows of Table 7.1. Transforming the boundary conditions from s to θ_f is simple:

$$\psi(s = 0) = \psi(\theta_f = 0),$$
$$\psi(s = \Gamma/4) = \psi(\theta_f = \pi - \arccos e),$$
$$\frac{\mathrm{d}}{\mathrm{d}s}\psi(s = 0) = \frac{1}{a - c}\frac{\mathrm{d}}{\mathrm{d}\theta_f}\psi(\theta_f = 0),$$
$$\frac{\mathrm{d}}{\mathrm{d}s}\psi(s = \Gamma/4) = \frac{\sqrt{1 - e^2}}{a}\frac{\mathrm{d}}{\mathrm{d}\theta_f}\psi(\theta_f = \pi - \arccos e). \tag{7.110}$$

To solve for a mode and its eigenvalue, the norm squared of the boundary quantity that should be zero at $s = \Gamma/4$ is minimized with respect to complex k. This type of solution is a modification of the *shooting method* for ODE boundary value problems, as the problem is solved by starting at zero and "aiming" at the correct value of on the other boundary. The ψ ODE integration can be numerically difficult because many cycles must be integrated over, and Eq. (7.66) must be solved many times per cycle. In our implementation, round-off error occasionally becomes a problem when 16 digit precision is used.

The simpler method is to use the WKB approximation. (The BOA, when combined with the WKB approximation, can be called the BO-WKB method.) The fact that $U(s)$ is periodic with period $\Gamma/2$, and that it is an even symmetric function about the vertex points $s = \text{integer} \times \Gamma/4$ (or $\theta = \text{integer} \times \pi/2$), allows one to perform a Fourier cosine series expansion in s or θ. In the variable θ, the Fourier expansion using N_e terms is

$$U(\theta) \approx \sum_{j=0}^{N_e - 1} c_j \cos(2j\theta). \tag{7.111}$$

All of the complex coefficients c_j can be calculated at once by solving N_e linear equations, each of which sets the expansion of U equal to the exact (numerical) U at a different value of θ on $[0, \pi/2]$. No integrations are necessary to find the expansion of U, (we have chosen point matching over integration, as we have sometimes done in the methods discussed in Chapters 1 and 2). The formula (I.3) is used to obtain $\theta(\theta_f)$.

One auspicious fact is that dU/ds and $dU/d\theta$ are zero at the vertex points. The necessary boundary quantities in the WKB approximation are then

$$\psi(0) = \frac{A+B}{\sqrt{|U(0)|}},$$

$$\psi(\Gamma/4) = \frac{A}{\sqrt{|U(\Gamma/4)|}} e^{i \int_0^{\Gamma/4} U(s)\, ds} + \frac{B}{\sqrt{|U(\Gamma/4)|}} e^{-i \int_0^{\Gamma/4} U(s)\, ds},$$

$$\psi'(0) = \frac{i(A-B)U(0)}{\sqrt{|U(0)|}},$$

$$\psi'(\Gamma/4) = \frac{iAU(\Gamma/4)}{\sqrt{|U(\Gamma/4)|}} e^{i \int_0^{\Gamma/4} U(s)\, ds} - \frac{iBU(\Gamma/4)}{\sqrt{|U(\Gamma/4)|}} e^{-i \int_0^{\Gamma/4} U(s)\, ds}. \qquad (7.112)$$

Using Table 7.1, one can see that the quantization condition for complex k for the $(+,-)$ and $(-,+)$ modes is

$$\cos\left[\int_0^{\Gamma/4} U(s)\, ds \right] = 0. \qquad (7.113)$$

The k quantization condition for the $(+,+)$ and $(-,-)$ modes is

$$\sin\left[\int_0^{\Gamma/4} U(s)\, ds \right] = 0. \qquad (7.114)$$

The two, two-fold degeneracies implied by these quantization conditions is an approximation inherent in the current approach; none of the four parity states are expected to be exactly degenerate.

The integral $\int_0^{\Gamma/4} U(s)\, ds = \int_0^{\pi/2} U(\theta)(ds/d\theta)\, d\theta$ must be performed using the θ expansion (7.111). The derivative $ds/d\theta$ is calculated in Eq. (I.2).

7.5.2 BO-WKB Estimation of Resonance Width

Here we show that, for a moderately sized oval resonator with a deformation sufficient to produce directional emission, the BO-WKB method does well at predicting the wavefunction Ψ as well as the real part of the wavenumber, $\operatorname{Re} k$. Results for $\operatorname{Im} k$ are found to deviate somewhat from the result of the direct numerical calculation. The direct numerical calculation is a point-matching method using the 2D angular momentum basis (see the end of Sec. 1.2.1).

We examine a dielectric ellipse (surrounded by air) with the parameters

$$
\begin{array}{cc}
a = 1 \qquad b = \frac{9}{10} \qquad e = c = 0.43589 \\
\Gamma/4 = 1.49329 \qquad \theta_f(\theta = \pi/2) = 2.02182 \\
\frac{81}{100} \le R \le \frac{10}{9} \qquad \sup(R'(s))^2 = 0.10028 \\
n = 1.5 \qquad 63 < \operatorname{Re}(k) < 65
\end{array}
\tag{7.115}
$$

The pre-BOA validity condition (7.23) and the BOA validity conditions (7.35, 7.36) are met by at least one decimal order of magnitude.

Using Eq. 7.113, with $U(\theta)$ approximated by a Fourier cosine series of $N_e = 16$ terms, we find a $(-, +)$ mode with 3 nodes (4 antinodes) in $\Phi(\rho)$ and 34.5 nodes and antinodes in $\psi(s)$ on $[0, \Gamma/4)$, having $k = 64.142 - 0.00611i$. The direct calculation of this mode yields $k = 64.127 - 0.00792i$, and the intensity plot given by the direct calculation is shown in Fig. 7.3. The neighboring $(-, +)$ mode, with 2 Φ nodes and (4×36.5) ψ nodes, has a BO-WKB k value of $63.862 - 0.00013i$ and a directly calculated k value of $63.854 - 0.00027i$. The other neighbor, having 4 Φ nodes and (4×32.5) ψ nodes, fairs better, with $k_{\text{BO-WKB}} = 64.100 - 0.0546i$ and $k_{\text{direct}} = 64.085 - 0.0657i$. A comparison done with similar k on the less deformed $a = 1$, $b = 0.98$ ellipse resulted in much better agreement: $k_{\text{BO-WKB}} = 64.6024 - 0.0203i$ and $k_{\text{direct}} = 64.6018 - 0.0207i$.

In many of the numerical methods presented in this work, there is a remarkably weak coupling between the real and imaginary parts of k. For example, changing the number of basis modes in the expansion of a dome cavity problem may change k from $100.00 - 4.1200 \times 10^{-6}$ to $100.01 - 4.1203 \times 10^{-6}$. If k, taken as a single complex number, were to undergo a random relative fluctuation of 0.01%, the imaginary part would loose all of its information. This is not what happened in the fictitious but typical case given here, the imaginary part of k receives a kick proportional to its

Figure 7.3: Intensity plot of ellipse mode described in text. A darker shade indicates greater intensity.
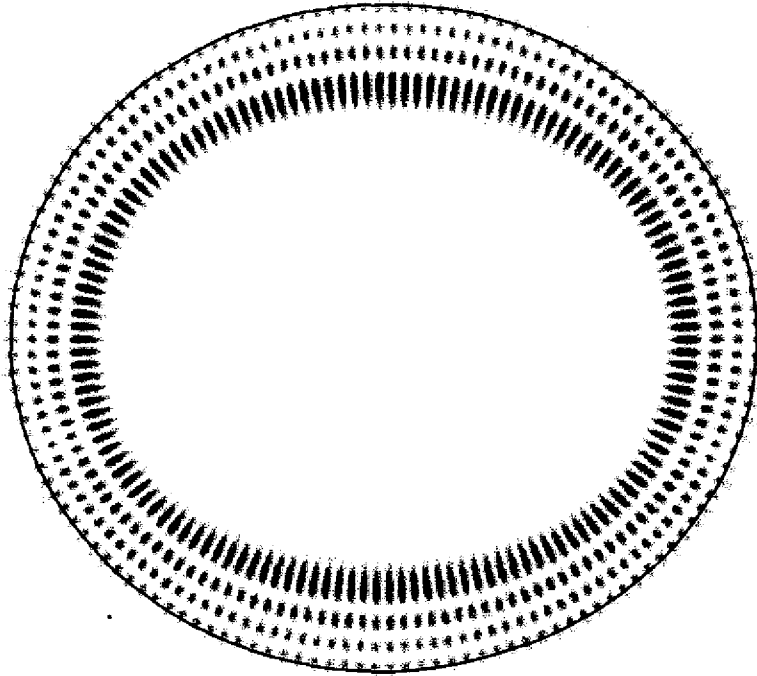
own size, and thus extremely small resonance widths may be calculated without being destroyed by numerical error. This "weak coupling" between $\mathrm{Re}\,k$ and $\mathrm{Im}\,k$ appears to be present in the direct numerical calculation of the WG modes. Unfortunately it appears that the BO-WKB method of finding k suffers from a significantly stronger coupling, although numerical fluctuations in k still do not behave as if k receives a random complex kick. The use of the the BO-WKB method to calculate resonance widths, is then limited to modes with sufficiently low Q factors. These modes have relatively low average angular momentum. In some ways this limitation of the BO-WKB method is surprising. The mental picture of a wave traveling along the edge of the resonator, loosing amplitude to transmission but gaining the same amount (on average) due to a fictitious gain embodied by $\mathrm{Im}\,k$, suggests that the imaginary part of k should be substantially independent of real part of k.

Perhaps the real surprise is that angular momentum basis expansion methods do so well for calculating resonance widths. It is not obvious why $\mathrm{Im}\,k$ and $\mathrm{Re}\,k$ should "decouple" in these methods.

7.5.3   Comparing Loss from $\mathrm{Im}\,U$ and Gain from $\mathrm{Im}\,k$

Here we look at a single example in which we compare the loss experienced by ψ in the waveguide picture with the "gain" experienced by the classical trajectory using the complex k which comes from the direct quasimode calculation.[6] We compare the LHS of Eq. (7.95) with the RHS. We divide this equation by 4 and state

$$\text{LHS} = \int_0^{\pi/2} \mathrm{Im}[U^{(2)}(\theta)]\frac{\mathrm{d}s}{\mathrm{d}\theta}\,\mathrm{d}\theta, \tag{7.116}$$

$$\text{RHS} = -\frac{n}{4}\mathrm{Im}(k)Z, \tag{7.117}$$

where Z is the length of the classical trajectory associated with one trip around the billiard/resonator, and k is the complex quasimode wavenumber found by the direct (numerical point-matching) calculation. To determine Z, a the initial conditions of a trajectory are picked, starting at some θ_0, with the ray having an angle of incidence χ_0 given by

$$\sin\chi_0 = \frac{\mathrm{Re}\,U^{(1)}(\theta_0)}{n\mathrm{Re}k}. \tag{7.118}$$

The classical trajectory is for many bounces, and the estimation of Z is found by finding a bounce event where $\theta \approx \theta_0$, and then dividing the total path length by the number of times the trajectory circled the origin.

We tested the relation for the quasimode mentioned in Sec. 7.5.2 having 4 Φ nodes, (4×32.5) ψ nodes, and $k = 64.085 - 0.0657i$. For this mode, LHS = 0.10901. Three estimations of Z were taken, one each for $\theta_0 = 0, \pi/4, \pi/2$. The resulting values of the RHS were 0.1107, 0.1103, and 0.1100, with a mean of 0.1103. The relative error using the mean RHS is 1.2%.

The result here, which of course should be joined by other calculations for non-circular oval resonators[7], suggests that the gain of the classical trajectory truly corresponds with the loss of the mode due to curvature radiation/tunneling. This correspondence has already been established due to the success of the sequential tunneling

---

[6]This statement would be more symmetric if the last phrase were "which comes from the resonator picture". However, since the BO-WKB has not produced very accurate estimations of $\mathrm{Im}\,k$ for the cases that we have examined, we use the k from the direct point-matching calculation instead of the $k^{(1)}$ of the BO-WKB method.

[7]For circular resonators, we already have the five data points of Table 7.2, which show reasonable agreement.

model developed by Nöckel [75] as discussed in Sec. 5.2. The BO-WKB method is in a sense the continuous version of the sequential tunneling model. Both BO-WKB and sequential tunneling use gain along the classical trajectory to balance losses incurred at the surface. In the sequential tunneling model the losses occur periodically (at the bounces), while in the BO-WKB method, the loss occurs continuously as the WG wave travels.

7.6 Conclusions for the BO-WKB method

We have extended the application of the generalized Born-Oppenheimer approximation from whispering gallery modes in metallic cavities [89, 90] to whispering gallery modes in dielectric cavities. The conditions for validity have been determined, as well as the validity of the WKB approximation, the use of which follows as a natural consequence. We have implemented the resulting BO-WKB method for the ellipse of eccentricity 0.436, and have found that the prediction of the mode structure and real wavenumber are accurate, but that the prediction of the small resonance width suffers significant error.

Additionally, our investigation has led to a better understanding of the relation of the gain quantity, $\mathrm{Im}\,k$, from the quasimode calculation and the resonator picture, to the loss quantity, $\mathrm{Im}\,U$, from the waveguide picture. The average ratio of the two quantities has been shown to be proportional to the average derivative $d(UR)/d(kR)$. The phase advance along the *caustic* due to $n\mathrm{Re}\,k^{(1)}$ is the same as the phase advance of the boundary wave $\psi(s)$, while the "gain" along the *classical trajectory* due to $n\mathrm{Im}\,k^{(1)}$ appears to be offset by the "loss" of the boundary wave $\psi(s)$ in the waveguide picture. The loss-gain equality in the BO-WKB procedure is a continuous version of the loss-gain equality behind the sequential tunneling model. The BO-WKB procedure is essentially a more advanced version of the sequential tunneling model; sequential tunneling is classical method augmented by wave behavior, the BO-WKB method lies closer to, if not in the realm of semiclassical techniques, as it is in part obtained by taking the short wavelength limit of actual wave equations.

The question of whether the BO-WKB method is practical is still open. At very large k values, the direct point-matching calculation of the modes may become intolerably slow. In this case the BO-WKB method may be the most efficient way to determine the whispering gallery quasimodes.

Chapter 8

CONCLUSIONS

For the dome cavity project, we have analysed how the true electromagnetic eigenmodes are chosen in the presense of a dielectric mirror, which produces nonlocal boundary conditions. To prepare for this analysis, we have developed a numerical program which solves the problem in two different bases, and which fully implements the scalar version of the problem as well as the vector version. To use the vector multipole basis with the plane wave reflection boundary conditions, we have developed full conversions between multipoles and plane waves of arbitrary angle.[1]

The dome cavity simulations have yielded two new results, paraxial mode mixing (or classical spin-orbit coupling) and the V modes. The former is of considerable importance both practically and theoretically. Nearly every optics table on the planet has a two-mirror cavity capped by dielectric mirrors (most lasers use these cavities). As cavities get smaller and mirrors become more reflective, researchers and engineers will increasingly find themselves in the regime in which the nominal paraxial degeneracy is observably lifted. This lifting was carefully analysed in metallic microwave cavities through the somewhat forgotten work of Erickson, Cullen, Yu and Luk [36, 49, 44, 45, 46, 47]. As optical cavities with dielectric mirrors venture into this regime, the paraxial mode mixing will suddenly become important, perhaps initially as an nuisance, but inevitably as a feature. Microlasers which output pure Laguerre-Gauss and mixed Laguerre-Gauss modes, and fiber filters which select and separate them, will undoubtedly find applications in nanomanipulation and communication. The fine mode splitting may also be useful in CQED schemes for quantum information processing.

---

[1]Similar results undoubtedly exist in older literature related to Mie scattering, but our derivation is "from scratch".

From the theoretical viewpoint, we have a tantalizing, difficult problem. Currently we are confronted with an apparent spin-orbit coupling with no Hamiltonian, no Schrödinger equation, no single-particle Dirac equation. Electromagnetic perturbation theory is currently limited in scope and in use (a Google search for "electromagnetic perturbation theory" scores about 30 hits, most of them irrelevant). Loosely speaking, the problem is vectorial, singular, nonlocal, degenerate, and perturbative. Through considerable effort, the author has put together an incomplete degenerate perturbation theory which remarkably reproduces approximately arctangent-shaped mixing curves. Through another conserable effort, Jens Nöckel has begun to create an effective theory using a generalized Born-Oppenheimer approximation in cavities. This theory has only been briefly mentioned in this work, due to newly discovered hurdles which occur at an early stage of the theory. A satisfactory two-state theory of mixing remains out of reach. Once discovered, it could perhaps shed light on other systems in which approximately degenerate families of modes are broken into modes having different structure.

The second part of this dissertation has focused on classical and semiclassical descriptions of circulating (whispering gallery) modes in 2D oval dielectric resonators. The sequential tunneling model, developed by Nöckel, yields good estimates of resonance widths for modes in circle (cylinder) resonators. In noncircular resonators, however, there are significant discrepancies between predicted and actual $\text{Im}\,k$. The logical extension of the sequential tunneling model is to include nonspecular effects, such as the Goos-Hänchen shift. Applied symmetrically, the GH shift has no effect in the circle resonator, making it a reasonable candidate to resolve the deformation-sensitive discrepancies. Unfortunately, we obtain the negative result that the GH shift is not well defined for the beam widths and lengths associated with whispering gallery modes. However, we do find a nontrivial and clean use of the GH effect in the dome cavity. The GH shift accurately predicts the character of the V modes. This is the only situation known to the author in which an incorporation of the GH shift into cavity ray dynamics has had an significant effect.

In a second approach to whispering gallery modes, we derive the BO-WKB method for oval dielectric resonators, starting from the Helmholtz equation. A similar series of approximations had previously been employed in conducting cavities. In addition to providing a potentially useful way to calculate WG modes and traveling waves, the BO-WKB study suggests that the classical orbits discussed previously in the context of sequential tunneling are indeed useful for describing mode decay. It is

found that, if the wave experiences a gain coefficient of $n|\mathrm{Im}\,k^{(1)}|$, while traveling along the classical path for one round trip, the amplitude increase is well matched to the round trip loss found by integrating $\mathrm{Im}\,U^{(2)}$ along the boundary. The values of both $\mathrm{Im}\,k^{(1)}$ and $\mathrm{Im}\,U^{(2)}(s)$ come from the BO-WKB procedure; quantities with (1) are calculated with "resonance picture" conditions and those with (2) are calculated with "waveguide picture" conditions. This matching using the classical path length is a result that one might have hoped for, but perhaps not really have expected—it gives a morale boost to classical models. More work will be necessary to thoroughly characterize the accuracy of this correspondence.

In future work, it would be worthwhile to investigate whether some sort of BOA can be made for scalar waves in 3D cavities. As part of an investigation with Scott Lacey, Hailin Wang, and Jens Nöckel [67] that is *not described in this work*, the author has developed a fast and extendable program to follow classical ray trajectories in 3D billiards. One interesting result is that a WG trajectory which visits the two polar regions in a prolate shape (ellipse or quadrupole) can be localized (stabilized) to a particular region of longitude (say the $0, 180°$ region) by a slight squishing of the shape in the orthogonal longitude (compression of the axis going through the equator and longitude $90, 270°$). The stabilizing deformation can be considerably (but not arbitrarily) smaller than the primary prolate deformation. What is seen is that the orbit approximately lies in a plane on intermediate timescales, and that this plane slowly precesses about the secondary, squished axis. Thus, the trajectory behaves, to some degree, as if it is in a 2D billiard which is adiabatically and periodically being perturbed. It was initially of interest to study this adiabatic problem to see if there existed usuable analogies to rigid bodies and torque. The adiabaticity is an interesting aspect. The theory of adiabatically varying chaotic systems has produced a rich and esoteric subfield [94, 95, 96, 97, 98, 29], which includes various effective forces caused by averaging over the fastest time scales and seeks important unifications of mechanics, statistical mechanics, and quantum chaos. It appears that the elegance and motivation associated with adiabatic billiard is lost when the billiard motion is not fully chaotic, and the adiabatic averaging was not pursued. There are however, interesting questions which come from thinking about this problem. For instance, if a static 3D billiard acts in some ways as an adiabatically varied 2D billiard, can a static 2D billiard be described in some sense by an adiabatically varied 1D billiard (a Fermi oscillator)? A comparison of SOS plots for a particular 1D Fermi oscillator (having a simple potential in addition to moving walls) and a particular 2D billiard is given in

Ref. [99]. The similarities in the structures are interesting, but no one has taken this beyond the "gee whiz" stage. The more we investigated the more differences we found between static $N + 1$ dimensional billiards and periodically perturbed N dimensional billiards, and useful results were not obtained. Still, it is possible that some articulable connection, some encompassing super-system, remains to be discovered.

Appendix A

EXPLICIT BESSEL WAVE EXPANSIONS OF LAGUERRE-GAUSS MODES

This appendix is an extension of the first portion of Sec. 3.1.4. The paraxial Bessel wave expansions given here derive from matching the electric field at the waist to the mode pattern $LG_p^l \sigma^s$, as in Eq. (3.19). The expansions cover unidirectional beams, standing waves formed by normal reflection of a beam off a planar mirror (generally a Bragg mirror), and actual cavity modes at the $\alpha = 0$ point on the mixing curve. A further use for these formulas, as the field contributions from the two basis modes from which mixed cavity modes are approximately formed, is the subject of Chap. 4. The expressions here are functions of z as well as ρ and ϕ.

We expand the beams/modes corresponding to the two mixable mode patterns of any mixable pair, $LG_{p_+}^{|l_+|} \sigma^+$ and $LG_{p_-}^{|l_-|} \sigma^-$, where

$$l_\pm = m \mp 1,$$
$$2p_+ + |l_+| = N,$$
$$2p_- + |l_-| = N. \tag{A.1}$$

(No beams/modes exist where $p_\pm$ is a half-integer.) The fields components used are E_+, E_-, E_z, H_+, H_-, and H_z, where the vector field for $\boldsymbol{E}$ or $\boldsymbol{H}$ is given by the form

$$\boldsymbol{V} = V_+ \sigma^+ + V_- \sigma^- + V_z \hat{z}. \tag{A.2}$$

A superscript "(+)" will be appended to field components corresponding to the $\mathrm{LG}_{p_+}^{|l_+|}\sigma^+$ beam/mode, and a superscript "(−)" will be appended for the $\mathrm{LG}_{p_-}^{|l_-|}\sigma^-$ beam/mode. The A mode pattern of Chap. 3 corresponds to a "+" mode, while the and B mode pattern corresponds to a "−" mode.

The expressions here use $\zeta_\pm$ from (3.28), as well as the definition

$$c(x) \equiv |\cos\theta_k|$$
$$= \sqrt{1 - \frac{2}{k^2 w_0^2}x^2} = \sqrt{1 - \frac{h^2}{2}x^2}, \tag{A.3}$$

(remember x is $(\sqrt{2}/h)\sin\theta_k$, not a spatial coordinate). For a beam traveling in the $+z$ direction, $s_u = p_u = 1$ and $s_d = p_d = 0$. For an incoming beam being reflected to form a standing wave, one can use $s_{u/d}$ and $p_{u/d}$ as defined in Eq. (3.20):

$$s_d = p_d = e^{ikz_1\cos\theta_k},$$
$$s_u = e^{-ikz_1\cos\theta_k}r_s(k,\theta_k),$$
$$p_u = e^{-ikz_1\cos\theta_k}r_p(k,\theta_k). \tag{A.4}$$

We assume a refractive index of 1, and, as in Sec. 3.1.4, the free-space waist of the mode must be at $z = 0$. Setting the origin to be at the waist means that the individual values of z_d and z_1 have been shifted while the length of the cavity, $L \equiv z_d - z_1$, remains constant and known.

Following the expansion of the electric and magnetic fields, Sec. A.3 discusses the leading order paraxial approximation. A discussion of z_1, relevant to the material here, is given in Sec. 4.1.1.

A.1 Electric Field

For the $\mathrm{LG}_{p_+}^{|l_+|}\sigma^+$ beam/mode, the electric field is given by:

$$E_+^{(+)}(x) = (-1)^{p_+}\frac{1}{\sqrt{2\pi}}\frac{\zeta_+}{w_0}\sqrt{\frac{p_+!}{(p_+ + |l_+|)!}}e^{il_+\phi}\int_0^{\sqrt{2}/h} dx\, x^{1+|l_+|}L_{p_+}^{|l_+|}(x^2)e^{-x^2/2}$$
$$\times J_{l_+}[\frac{\sqrt{2}}{w_0}\rho x]\left(\frac{p_u(x)e^{ikc(x)z} + p_d(x)e^{-ikc(x)z}}{p_u(x) + p_d(x)} + \frac{s_u(x)e^{ikc(x)z} + s_d(x)e^{-ikc(x)z}}{s_u(x) + s_d(x)}\right) \tag{A.5}$$

$$E_-^{(+)}(\boldsymbol{x}) = (-1)^{p_+} \frac{1}{\sqrt{2\pi}} \frac{\zeta_+}{w_0} \sqrt{\frac{p_+!}{(p_+ + |l_+|)!}} e^{il_-\phi} \int_0^{\sqrt{2}/h} \mathrm{d}x\, x^{1+|l_+|} L_{p_+}^{|l_+|}(x^2) e^{-x^2/2}$$

$$\times\, J_{l_-}[\frac{\sqrt{2}}{w_0}\rho x]\left(\frac{s_\mathrm{u}(x)e^{ikc(x)z} + s_\mathrm{d}(x)e^{-ikc(x)z}}{s_\mathrm{u}(x) + s_\mathrm{d}(x)} - \frac{p_\mathrm{u}(x)e^{ikc(x)z} + p_\mathrm{d}(x)e^{-ikc(x)z}}{p_\mathrm{u}(x) + p_\mathrm{d}(x)}\right)$$

$$\hspace{12cm}\text{(A.6)}$$

$$E_z^{(+)}(\boldsymbol{x}) = -i(-1)^{p_+} \frac{h}{\sqrt{2\pi}} \frac{\zeta_+}{w_0} \sqrt{\frac{p_+!}{(p_+ + |l_+|)!}} e^{im\phi} \int_0^{\sqrt{2}/h} \mathrm{d}x\, \frac{1}{c(x)} x^{2+|l_+|} L_{p_+}^{|l_+|}(x^2) e^{-x^2/2}$$

$$\times\, J_m[\frac{\sqrt{2}}{w_0}\rho x]\left(\frac{p_\mathrm{u}(x)e^{ikc(x)z} - p_\mathrm{d}(x)e^{-ikc(x)z}}{p_\mathrm{u}(x) + p_\mathrm{d}(x)}\right)$$

$$\hspace{12cm}\text{(A.7)}$$

Note that $E_-^{(+)}$ can be nonzero (except at $z = 0$) if $r_\mathrm{s} \neq r_\mathrm{p}$.

For the $\mathrm{LG}_{p_-}^{|l_-|}\sigma^-$ beam/mode, the electric field is given by:

$$E_+^{(-)} = (-1)^{p_-} \frac{1}{\sqrt{2\pi}} \frac{\zeta_-}{w_0} \sqrt{\frac{p_-!}{(p_- + |l_-|)!}} e^{il_+\phi} \int_0^{\sqrt{2}/h} \mathrm{d}x\, x^{1+|l_-|} L_{p_-}^{|l_-|}(x^2) e^{-x^2/2}$$

$$\times\, J_{l_+}[\frac{\sqrt{2}}{w_0}\rho x]\left(\frac{s_\mathrm{u}(x)e^{ikc(x)z} + s_\mathrm{d}(x)e^{-ikc(x)z}}{s_\mathrm{u}(x) + s_\mathrm{d}(x)} - \frac{p_\mathrm{u}(x)e^{ikc(x)z} + p_\mathrm{d}(x)e^{-ikc(x)z}}{p_\mathrm{u}(x) + p_\mathrm{d}(x)}\right)$$

$$\hspace{12cm}\text{(A.8)}$$

$$E_-^{(-)} = (-1)^{p_-} \frac{1}{\sqrt{2\pi}} \frac{\zeta_-}{w_0} \sqrt{\frac{p_-!}{(p_- + |l_-|)!}} e^{il_-\phi} \int_0^{\sqrt{2}/h} \mathrm{d}x\, x^{1+|l_-|} L_{p_-}^{|l_-|}(x^2) e^{-x^2/2}$$

$$\times\, J_{l_-}[\frac{\sqrt{2}}{w_0}\rho x]\left(\frac{p_\mathrm{u}(x)e^{ikc(x)z} + p_\mathrm{d}(x)e^{-ikc(x)z}}{p_\mathrm{u}(x) + p_\mathrm{d}(x)} + \frac{s_\mathrm{u}(x)e^{ikc(x)z} + s_\mathrm{d}(x)e^{-ikc(x)z}}{s_\mathrm{u}(x) + s_\mathrm{d}(x)}\right)$$

$$\hspace{12cm}\text{(A.9)}$$

$$E_z^{(-)} = i(-1)^{p_-} \frac{h}{\sqrt{2\pi}} \frac{\zeta_-}{w_0} \sqrt{\frac{p_-!}{(p_- + |l_-|)!}} e^{im\phi} \int_0^{\sqrt{2}/h} \mathrm{d}x\, \frac{1}{c(x)} x^{2+|l_-|} L_{p_-}^{|l_-|}(x^2) e^{-x^2/2}$$

$$\times\, J_m[\frac{\sqrt{2}}{w_0}\rho x]\left(\frac{p_\mathrm{u}(x)e^{ikc(x)z} - p_\mathrm{d}(x)e^{-ikc(x)z}}{p_\mathrm{u}(x) + p_\mathrm{d}(x)}\right)$$

$$\hspace{12cm}\text{(A.10)}$$

Again, if $r_s \neq r_p$ then $E_+^{(-)}$ can be nonzero (except at $z = 0$).

A.2 Magnetic Field

We must first redo the steps of Sec. 2.4.2 for the magnetic field $\boldsymbol{H}$. We divide $\boldsymbol{H}$ into $\boldsymbol{H}_{\mathrm{s}}$ and $\boldsymbol{H}_{\mathrm{p}}$ so that $\boldsymbol{H}_{\mathrm{s}} \cdot \hat{z} = 0$ and $\boldsymbol{H}_{\mathrm{p}} \cdot \hat{\phi} = 0$. The plane wave expansion is

$$\boldsymbol{H}_{\mathrm{s}} = -\int \mathrm{d}\Omega_k \, \mathrm{sgn}(\cos\theta_k) \tilde{P}_k \epsilon_{\mathrm{s},k} e^{i\boldsymbol{k}\cdot\boldsymbol{x}},$$

$$\boldsymbol{H}_{\mathrm{p}} = \int \mathrm{d}\Omega_k \, \mathrm{sgn}(\cos\theta_k) \tilde{S}_k \epsilon_{\mathrm{p},k} e^{i\boldsymbol{k}\cdot\boldsymbol{x}}. \tag{A.11}$$

Comparing with Eq. (2.17) shows that in Eqs. (2.19, 2.20, 2.21) one can make the symbol replacements: $E_* \to H_*$ (where $*$ is any symbol), $S_\pm \to S_\pm^{(H)}$, $P_\pm \to P_\pm^{(H)}$, $P_k \to \mathrm{sgn}(\cos\theta_k) S_k$ and $S_k \to -\mathrm{sgn}(\cos\theta_k) P_k$. These new expressions must then be combined with Eqs. (3.25) and (3.27).

For the $\mathrm{LG}_{p_+}^{|l_+|}\sigma^+$ beam/mode, the magnetic field is given by:

$$
\begin{aligned}
H_+^{(+)} = {}& -i(-1)^{p_+}\frac{1}{\sqrt{2\pi}}\frac{\zeta_+}{w_0}\sqrt{\frac{p_+!}{(p_+ + |l_+|)!}}e^{il_+\phi}\int_0^{\sqrt{2}/h}\mathrm{d}x\, x^{1+|l_+|}L_{p_+}^{|l_+|}(x^2)e^{-x^2/2} \\
& \times J_{l_+}[\frac{\sqrt{2}}{w_0}\rho x]\left(\frac{1}{c(x)}\frac{p_{\mathrm{u}}(x)e^{ikc(x)z} - p_{\mathrm{d}}(x)e^{-ikc(x)z}}{p_{\mathrm{u}}(x) + p_{\mathrm{d}}(x)}\right. \\
& \left. + c(x)\frac{s_{\mathrm{u}}(x)e^{ikc(x)z} - s_{\mathrm{d}}(x)e^{-ikc(x)z}}{s_{\mathrm{u}}(x) + s_{\mathrm{d}}(x)}\right)
\end{aligned}
\tag{A.12}
$$

$$
\begin{aligned}
H_-^{(+)} = {}& -i(-1)^{p_+}\frac{1}{\sqrt{2\pi}}\frac{\zeta_+}{w_0}\sqrt{\frac{p_+!}{(p_+ + |l_+|)!}}e^{il_-\phi}\int_0^{\sqrt{2}/h}\mathrm{d}x\, x^{1+|l_+|}L_{p_+}^{|l_+|}(x^2)e^{-x^2/2} \\
& \times J_{l_-}[\frac{\sqrt{2}}{w_0}\rho x]\left(\frac{1}{c(x)}\frac{p_{\mathrm{u}}(x)e^{ikc(x)z} - p_{\mathrm{d}}(x)e^{-ikc(x)z}}{p_{\mathrm{u}}(x) + p_{\mathrm{d}}(x)}\right. \\
& \left. - c(x)\frac{s_{\mathrm{u}}(x)e^{ikc(x)z} - s_{\mathrm{d}}(x)e^{-ikc(x)z}}{s_{\mathrm{u}}(x) + s_{\mathrm{d}}(x)}\right)
\end{aligned}
\tag{A.13}
$$

$$
\begin{aligned}
H_z^{(+)} = {}& -(-1)^{p_+}\frac{h}{\sqrt{2\pi}}\frac{\zeta_+}{w_0}\sqrt{\frac{p_+!}{(p_+ + |l_+|)!}}e^{im\phi}\int_0^{\sqrt{2}/h}\mathrm{d}x\, x^{2+|l_+|}L_{p_+}^{|l_+|}(x^2)e^{-x^2/2} \\
& \times J_m[\frac{\sqrt{2}}{w_0}\rho x]\left(\frac{s_{\mathrm{u}}(x)e^{ikc(x)z} + s_{\mathrm{d}}(x)e^{-ikc(x)z}}{s_{\mathrm{u}}(x) + s_{\mathrm{d}}(x)}\right)
\end{aligned}
\tag{A.14}
$$

For the $LG_{p_-}^{|l_-|}\sigma^-$ beam/mode, the magnetic field is given by:

$$H_+^{(-)} = i(-1)^{p_-} \frac{1}{\sqrt{2\pi}} \frac{\zeta_-}{w_0} \sqrt{\frac{p_-!}{(p_- + |l_-|)!}} e^{il_+ \phi} \int_0^{\sqrt{2}/h} dx\, x^{1+|l_-|} L_{p_-}^{|l_-|}(x^2) e^{-x^2/2}$$

$$\times\, J_{l_+}[\frac{\sqrt{2}}{w_0}\rho x]\left(\frac{1}{c(x)} \frac{p_u(x)e^{ikc(x)z} - p_d(x)e^{-ikc(x)z}}{p_u(x) + p_d(x)}\right.$$

$$\left. - c(x)\frac{s_u(x)e^{ikc(x)z} - s_d(x)e^{-ikc(x)z}}{s_u(x) + s_d(x)}\right) \tag{A.15}$$

$$H_-^{(-)} = i(-1)^{p_-} \frac{1}{\sqrt{2\pi}} \frac{\zeta_-}{w_0} \sqrt{\frac{p_-!}{(p_- + |l_-|)!}} e^{il_- \phi} \int_0^{\sqrt{2}/h} dx\, x^{1+|l_-|} L_{p_-}^{|l_-|}(x^2) e^{-x^2/2}$$

$$\times\, J_{l_-}[\frac{\sqrt{2}}{w_0}\rho x]\left(\frac{1}{c(x)} \frac{p_u(x)e^{ikc(x)z} - p_d(x)e^{-ikc(x)z}}{p_u(x) + p_d(x)}\right.$$

$$\left. + c(x)\frac{s_u(x)e^{ikc(x)z} - s_d(x)e^{-ikc(x)z}}{s_u(x) + s_d(x)}\right) \tag{A.16}$$

$$H_z^{(-)} = -(-1)^{p_-} \frac{h}{\sqrt{2\pi}} \frac{\zeta_-}{w_0} \sqrt{\frac{p_-!}{(p_- + |l_-|)!}} e^{im\phi} \int_0^{\sqrt{2}/h} dx\, x^{2+|l_-|} L_{p_-}^{|l_-|}(x^2) e^{-x^2/2}$$

$$\times\, J_m[\frac{\sqrt{2}}{w_0}\rho x]\left(\frac{s_u(x)e^{ikc(x)z} + s_d(x)e^{-ikc(x)z}}{s_u(x) + s_d(x)}\right) \tag{A.17}$$

A.3 Paraxial Expansions of the Fields

The end of Sec. 3.1.4 shows that if the Bessel wave form is kept to $O(\theta_k^2)$, which here translates to $O(h^2)$, the paraxial wave equation is solved exactly. If higher orders of h are kept in the Bessel wave, the expansion would presumably satisfy a higher order paraxial wave equation that one could generate from the discussions in Lax *et al.* [39]. However, one does not consistently gain accuracy unless the expansion itself has this higher order accuracy. Since the expansions in this section are designed to give the spatial $LG\sigma$ patterns at $z = 0$, and these patterns came from the regular paraxial wave equation, we cannot expect to at this point gain accuracy, say by approximating $c(x)$ as $1 - (h^2/4)x^2 - (h^4/32)x^4$ instead of just $1 - (h^2/4)x^2$. It seems that the consistent approach is to expand the quantities that are shown to be in the integrands in Eqs. (A.5-A.10, A.12-A.17) to second order in h. This means, for example, expanding r_s, which is found in s_u, as $e^{i\phi_0}(1 + i\epsilon_s h^2 x^2/2)$ (here we are assuming $r_{s/p}$ to be of unit modulus).

The expansions are straightforward. One interesting result is that the quanti-

ties $E_{\pm}^{\mp}$ and $H_{\pm}^{\mp}$ are of leading order h^2. Explicitly, the quantity within the large parentheses in Eq. (A.6) is

$$(\epsilon_s - \epsilon_p)h^2 x^2 \sin(kz)\frac{-e^{i(\phi_0 - 2kz_1)}}{(1 + e^{i(\phi_0 - 2kz_1)})^2}. \tag{A.18}$$

The presence of the quantity $\epsilon_s - \epsilon_p$ is intriguing, as this quantity appears to be the most universal predictor of mixing angle (see Fig. 3.11).

Appendix B

SPIN, ORBITAL, AND TOTAL ANGULAR MOMENTUM OF PARAXIAL BEAMS AND MODES

In Sec. 2.3 it was shown that, due to cylindrical symmetry, there exists a good quantum number m. From Noether's theorem we know that if a Lagrangian system is invariant under translations in ϕ (axial rotations), there exists a constant of the motion we may call the angular momentum. From these two facts we may guess that m has something to do with the total angular momentum of the field. Our approach to exploring the angular momentum will not involve field Lagrangians, however. In this appendix we explicitly show that the physical angular momentum of the electromagnetic field modes with quantum number m is indeed $m\hbar$ per quantum excitation. We also show that the volume-integrated spin is $s\hbar$ per photon, and hence the subtraction of spin from total angular momentum yields $l\hbar$, which is the orbital angular momentum (OAM). Since its discovery in 1992 by Allen et al. [22], there has been a great deal of theoretical and experimental work done on the OAM of beams, including discussions of generation [100, 101, 102], manipulation applications (optical tweezers) [52, 103, 53], quantum and classical information applications [56, 58, 104], and theoretical representations [105, 106, 107]. For more direct approaches to the separation of the spin and orbital angular momentum in beams, the reader is referred to Refs. [22, 108, 109]. A discussion of the extent to which orbital and spin angu-

lar momentum operators can be constructed for electromagnetic beams is given in Ref. [107]. Discussions of the forces and torques exerted on atoms, particles, and dielectric layers by beams carrying angular momentum are given in Refs. [109, 110, 107], while Ref. [102] describes the creation of an LG beam by coherently transferring angular momentum from a group atoms undergoing Larmor precession. Reviews of the orbital angular momentum of light include Refs. [23, 111].

B.1 Total Angular Momentum

Here we explicitly show that the quantized total angular momentum of a mixed cavity mode given by

$$\Psi = \cos(\alpha)\mathrm{LG}_p^{m-1}\sigma^+ - \sin(\alpha)\mathrm{LG}_{p-1}^{m+1}\sigma^-$$ (B.1)

is $m\hbar$. The equivalence of physical angular momentum and m has been shown, via an explicit calculation similar to the one that follows, by Barnett and Allen [108] for *arbitrary* paraxial Bessel-expanded beams with a certain ϕ dependence (E_x, $E_y \propto e^{il'\phi}$). However, we choose the specific form Ψ to show how this property works out for our paraxial *modes*, and so that the calculation of the spin in the next section is trivial. The basic strategy for calculating the quantized angular momentum is to show that the ratio of J_z, the classical angular momentum of the field, to U, the classical energy of the field, is of the form $\frac{m'}{\omega}$, where m' is an integer. Using the fact that light comes in quantized energy units of $\hbar\omega$, the quantized momentum is $\frac{m'}{\omega} \times \hbar\omega = \hbar m'$.

To show that the constants come out correctly, we will include ϵ_0, μ_0, and c. The wavenumber k is assumed to be real and free space will be considered for simplicity. We will always be working in the paraxial limit $h \ll 1$. (Barnett and Allen show that, outside the paraxial limit, separation of total angular momentum into spin and OAM is not defined.)

B.1.1 Classical Energy of a Mode

The physical field energy in a volume with complex field quantities is

$$U = \frac{\epsilon_0}{2} \int \mathrm{d}V\, \boldsymbol{E} \cdot \boldsymbol{E}^*.$$ (B.2)

The complex square of E is given by

$$\boldsymbol{E} \cdot \boldsymbol{E}^* = |E_+|^2 + |E_-|^2 + |E_z|^2. \tag{B.3}$$

For a "+" mode (such as mode A) the first term is order h^0, the second term is order h^4, and the third term is order h^2. For a "−" mode, the orders of the first and second terms are transposed. Only the $O(h^0)$ term, $|E_+^{(+)}|^2$ or $|E_-^{(-)}|^2$, need be kept. Here the superscript denotes whether the mode is "+" or "−". Using the expressions from Appendix A and keeping only $O(h^0)$ terms yields

$$U = \frac{\epsilon_0}{2} \int dV \left[\cos^2(\alpha)|E_+^{(+)}|^2 + \sin^2(\alpha)|E_-^{(-)}|^2 \right.$$
$$\left. - \cos(\alpha)\sin(\alpha)\left(\boldsymbol{E}_+^{(+)} \cdot \boldsymbol{E}_-^{(-)*} + \boldsymbol{E}_-^{(-)} \cdot \boldsymbol{E}_+^{(+)*} \right) \right]. \tag{B.4}$$

Performing the ϕ integral destroys the cross terms, while yielding a factor of 2π for the surviving terms. The result is

$$U = \epsilon_0 \pi \int_{z_a}^{z_b} dz \int_0^\infty d\rho\, \rho \int_0^\infty dx\, x \int_0^\infty dx'\, x' \left| 2 \frac{e^{i[\phi_0/2 + k(z-z_1)]} + e^{-i[\phi_0/2 + k(z-z_1)]}}{e^{i[\phi_0/2 - kz_1]} + e^{-i[\phi_0/2 - kz_1]}} \right|^2$$
$$\times \left[\cos^2(\alpha)\eta_+^2 f_+(x) f_+(x') J_{l_+}(\frac{\sqrt{2}}{w_0}\rho x) J_{l_+}(\frac{\sqrt{2}}{w_0}\rho x') \right.$$
$$\left. + \sin^2(\alpha)\eta_-^2 f_-(x) f_-(x') J_{l_-}(\frac{\sqrt{2}}{w_0}\rho x) J_{l_-}(\frac{\sqrt{2}}{w_0}\rho x') \right], \tag{B.5}$$

where

$$\eta_\pm \equiv \frac{(-1)^{p_\pm}\zeta_\pm}{\sqrt{2\pi}w_0} \sqrt{\frac{p_\pm!}{(p_\pm + |l_\pm|)!}}, \tag{B.6}$$

$$f_\pm(x) \equiv x^{|l_\pm|} L_{p_\pm}^{|l_\pm|}(x^2) e^{-x^2/2} \tag{B.7}$$

and the z integration is over some length, $\Delta z = z_b - z_a$, which can be the cavity length or a single wavelength. We will assume the contributions of oscillating functions of z

to the z integral are negligible. The expression then becomes

$$
U = \frac{\epsilon_0 2\pi \Delta z}{\cos^2(\phi_0/2 - kz_1)} \int_0^\infty \mathrm{d}\rho\, \rho \int_0^\infty \mathrm{d}x\, x \int_0^\infty \mathrm{d}x'\, x'
$$
$$
\times \left[\cos^2(\alpha)\eta_+^2 f_+(x) f_+(x') J_{l_+}(\frac{\sqrt{2}}{w_0}\rho x) J_{l_+}(\frac{\sqrt{2}}{w_0}\rho x') \right.
$$
$$
\left. + \sin^2(\alpha)\eta_-^2 f_-(x) f_-(x') J_{l_-}(\frac{\sqrt{2}}{w_0}\rho x) J_{l_-}(\frac{\sqrt{2}}{w_0}\rho x') \right]. \tag{B.8}
$$

Using the Bessel function orthogonality relation

$$
\int_0^\infty \mathrm{d}\rho\, \rho J_n(y\rho) J_n(y'\rho) = \frac{1}{y}\delta(y' - y), \tag{B.9}
$$

the result becomes

$$
U = \frac{\epsilon_0 \pi \Delta z w_0^2}{\cos^2(\phi_0/2 - kz_1)} \int_0^\infty \mathrm{d}x\, x \left[\cos^2(\alpha)\eta_+^2 f_+^2(x) + \sin^2(\alpha)\eta_-^2 f_-^2(x) \right]. \tag{B.10}
$$

The normalization relation

$$
\int_0^\infty \mathrm{d}x\, x f_\pm^2(x) = \frac{1}{2}\frac{(p_\pm + |l_\pm|)!}{p_\pm!}
$$
$$
= \frac{1}{4\pi w_0^2 \eta_\pm^2}, \tag{B.11}
$$

can be derived from orthogonality integral 7.414.3 of Ref. [48]:

$$
\int_0^\infty e^{-x} x^k L_n^k(x) L_m^k(x)\, \mathrm{d}x = \delta_{mn}\frac{(n+k)!}{n!}. \tag{B.12}
$$

Using (B.11) in (B.10) at last yields the simple result

$$
U = \frac{\epsilon_0}{4}\frac{\Delta z}{\cos^2(\phi_0/2 - kz_1)}. \tag{B.13}
$$

B.1.2 Classical Angular Momentum of a Mode

The classical angular momentum density of the (complex) electromagnetic field is given by

$$
\boldsymbol{j} = \epsilon_0\mu_0 \boldsymbol{r} \times (\boldsymbol{E} \times \boldsymbol{H}^*), \tag{B.14}
$$

while the total angular momentum is simply j_z integrated over volume. Performing the vector product with circular polarization components and taking the z component yields

$$j_z = \frac{1}{2\sqrt{2}c^2}\rho[e^{-i\phi}E_z H_+^* + e^{i\phi}E_z H_-^* - e^{i\phi}E_+ H_z^* - e^{-i\phi}E_- H_z^*] \qquad (B.15)$$

Each term is independent of ϕ, so that taking the ϕ integral yields

$$\int_0^{2\pi} j_z \mathrm{d}\phi = \frac{\pi}{\sqrt{2}c^2}\rho[E_z H_+^* + E_z H_-^* - E_+ H_z^* - E_- H_z^*]_{\phi=0}. \qquad (B.16)$$

Inserting the mixed mode $\boldsymbol{\Psi}$ and the dropping $O(h^2)$ fields $E_\pm^{(\mp)}$ and $H_\pm^{(\mp)}$ gives the total angular momentum

$$\begin{aligned}
J_z = \frac{\pi}{\sqrt{2}c^2}\int_{z_a}^{z_b}\mathrm{d}z\int_0^\infty \mathrm{d}\rho\,\rho^2 & \Big[\cos^2(\alpha)(E_z^{(+)}H_+^{(+)*} - E_+^{(+)}H_z^{(+)*}) \\
& + \sin^2(\alpha)(E_z^{(-)}H_-^{(-)*} - E_-^{(-)}H_z^{(-)*}) \\
& + \sin(\alpha)\cos(\alpha)(-E_z^{(+)}H_-^{(-)*} - E_z^{(-)}H_+^{(+)*} + E_-^{(-)}H_z^{(+)*} + E_+^{(+)}H_z^{(-)*})\Big]_{\phi=0} \quad (B.17)
\end{aligned}$$

Substituting from Appendix A to lowest order in h gives

$$E_\pm^{(\pm)}\Big|_{\phi=0} = 2\eta_\pm \frac{\cos[k(z-z_1)+\phi_0/2]}{\cos(\phi_0/2 - kz_1)}\int_0^\infty \mathrm{d}x'\,x'f_\pm(x')J_{l_\pm}(\frac{\sqrt{2}}{w_0}\rho x')$$

$$E_z^{(\pm)}\Big|_{\phi=0} = \pm h\eta_\pm \frac{\sin[k(z-z_1)+\phi_0/2]}{\cos(\phi_0/2 - kz_1)}\int_0^\infty \mathrm{d}x\,x^2 f_\pm(x)J_m(\frac{\sqrt{2}}{w_0}\rho x)$$

$$H_\pm^{(\pm)}\Big|_{\phi=0} = \pm 2\sqrt{\frac{\epsilon_0}{\mu_0}}\eta_\pm \frac{\sin[k(z-z_1)+\phi_0/2]}{\cos(\phi_0/2 - kz_1)}\int_0^\infty \mathrm{d}x'\,x'f_\pm(x')J_{l_\pm}(\frac{\sqrt{2}}{w_0}\rho x')$$

$$H_z^{(\pm)}\Big|_{\phi=0} = -h\sqrt{\frac{\epsilon_0}{\mu_0}}\eta_\pm \frac{\cos[k(z-z_1)+\phi_0/2]}{\cos(\phi_0/2 - kz_1)}\int_0^\infty \mathrm{d}x\,x^2 f_\pm(x)J_m(\frac{\sqrt{2}}{w_0}\rho x). \qquad (B.18)$$

Note that these quantities are real, so the complex conjugate can be dropped from the notation. One can see that each of the products in Eq. (B.17) has a factor of either $\cos^2[k(z-z_1)+\phi_0/2]$ or $\sin^2[k(z-z_1)+\phi_0/2]$. Performing the z integration transforms these factors into $\Delta z/2$. At this point one can see that $E_+^{(+)}H_z^{(-)} = E_z^{(-)}H_+^{(+)}$ and $E_-^{(-)}H_z^{(+)} = E_z^{(+)}H_-^{(-)}$ so that the term in Eq. (B.17) proportional to $\sin(\alpha)\cos(\alpha)$ is zero. One can also see that $E_+^{(+)}H_z^{(+)} = -E_z^{(+)}H_+^{(+)}$ and $E_-^{(-)}H_z^{(-)} = -E_z^{(-)}H_-^{(-)}$ which simplifies the other two terms.

The expression for J_z has now been simplified to

$$J_z = \frac{\epsilon_0 \pi \sqrt{2} h \Delta z}{c \cos^2(\phi_0/2 - kz_1)}$$
$$\times \left[\cos^2(\alpha)\eta_+^2 \int_0^\infty \mathrm{d}x\, x^2 f_+(x) \int_0^\infty \mathrm{d}x'\, x' f_+(x') I_+ \right.$$
$$\left. + \sin^2(\alpha)\eta_-^2 \int_0^\infty \mathrm{d}x\, x^2 f_-(x) \int_0^\infty \mathrm{d}x'\, x' f_-(x') I_- \right], \qquad \text{(B.19)}$$

where

$$I_\pm = \int_0^\infty \rho^2 J_m\left(\frac{\sqrt{2}}{w_0}x\rho\right) J_{m\mp1}\left(\frac{\sqrt{2}}{w_0}x'\rho\right) \mathrm{d}\rho. \qquad \text{(B.20)}$$

Using the Bessel function relation

$$J_{n\mp1}(y) = \frac{n}{y}J_n(y) \pm \frac{\mathrm{d}}{\mathrm{d}y}J_n(y), \qquad \text{(B.21)}$$

yields

$$I_\pm = \int_0^\infty \rho^2 J_m\left(\frac{\sqrt{2}}{w_0}x\rho\right) \left[\frac{m}{\frac{\sqrt{2}}{w_0}x'\rho}J_m\left(\frac{\sqrt{2}}{w_0}x'\rho\right) \pm J_m'\left(\frac{\sqrt{2}}{w_0}x'\rho\right)\right] \mathrm{d}\rho$$

$$= \frac{mw_0}{\sqrt{2}x'} \int_0^\infty \rho J_m\left(\frac{\sqrt{2}}{w_0}x\rho\right) J_m\left(\frac{\sqrt{2}}{w_0}x'\rho\right) \mathrm{d}\rho$$

$$\quad + \int_0^\infty \rho^2 J_m\left(\frac{\sqrt{2}}{w_0}x\rho\right) \frac{1}{\frac{\sqrt{2}}{w_0}\rho}\frac{\mathrm{d}}{\mathrm{d}x'}J_m\left(\frac{\sqrt{2}}{w_0}x'\rho\right) \mathrm{d}\rho$$

$$= \frac{mw_0^3}{2\sqrt{2}}x'^2\delta(x'-x) \pm \frac{w_0^3}{2\sqrt{2}}\frac{\mathrm{d}}{\mathrm{d}x'}\left(\frac{1}{x'}\delta(x'-x)\right)$$

$$= \frac{(m\mp1)w_0^3}{2\sqrt{2}x'^2}\delta(x'-x) \pm \frac{w_0^3}{2\sqrt{2}x'}\frac{\mathrm{d}}{\mathrm{d}x'}\delta(x'-x), \qquad \text{(B.22)}$$

where the derivative of the Bessel function w.r.t. its argument has been chosen to be evaluated via the derivative w.r.t. x' in order to take this derivative outside of the ρ integral, which is then evaluated using the orthogonality relation (B.9).

Performing the x' and x integrals of Eq. (B.19) on first term of $I_\pm$ is trivial via

Eq. (B.11), yielding

$$\frac{(m \mp 1)w_0}{8\sqrt{2\pi}\eta_{\pm}^2}. \tag{B.23}$$

The double integration of the second term of $I_{\pm}$ is

$$K_{\pm} \equiv \pm\frac{w_0^3}{2\sqrt{2}} \int_0^\infty \mathrm{d}x\, x^2 f_{\pm}(x) \int_0^\infty \mathrm{d}x'\, f_{\pm}(x')\frac{\mathrm{d}}{\mathrm{d}x'}\delta(x'-x)$$

$$= \mp\frac{w_0^3}{2\sqrt{2}} \int_0^\infty \mathrm{d}x\, x^2 f_{\pm}(x)f'_{\pm}(x). \tag{B.24}$$

Integrating by parts yields a boundary term of zero and leads to

$$K_{\pm} = \pm 2\left(\frac{w_0^3}{2\sqrt{2}}\right) \int_0^\infty x f_{\pm}^2(x)\, \mathrm{d}x - K_{\pm}$$

$$= \pm\left(\frac{w_0^3}{2\sqrt{2}}\right) \int_0^\infty x f_{\pm}^2(x)\, \mathrm{d}x$$

$$= \pm\frac{w_0}{8\sqrt{2\pi}\eta_{\pm}^2}. \tag{B.25}$$

Thus, with the result of the first term (B.23) we obtain

$$\int_0^\infty \mathrm{d}x\, x^2 f_{\pm}(x) \int_0^\infty \mathrm{d}x'\, x' f_{\pm}(x')I_{\pm} = \frac{mw_0}{8\sqrt{2\pi}\eta_{\pm}^2}. \tag{B.26}$$

Substituting this into Eq. (B.19) and using $h = 2/(kw_0)$ and $\omega = ck$ finally yields the total angular momentum

$$J_z = \frac{m\epsilon_0\Delta z}{4\omega\cos^2(\phi_0/2 - kz_1)}. \tag{B.27}$$

Comparing with the total energy (B.13) gives the expected ratio

$$\frac{J_z}{U} = \frac{m}{\omega}. \tag{B.28}$$

Thus by the argument at the beginning of Sec. B.1, the quantized momentum is $m\hbar$.

B.2 Spin and Orbital Angular Momentum

There is no clear separation of a spin term and an orbital term in the derivation of J_z in the previous section. Methods which have clear separation of orbital and angular momentum will briefly be mentioned at the end of this section. Here we give a brief physical argument for the spin-orbit separation via a definition of spin angular momentum density from the perspective of semiclassical quantum optics, in which the field is classical and the atoms are quantum mechanical.

Since an atom is small, we can only expect the intrinsic angular momentum, or spin, of light to be exchanged with the state of the atom. The extrinsic angular momentum[1], due to azimuthal field gradients, should not affect the internal state of the atom. Semiclassically, a locally σ^- polarized classical electromagnetic field (laser pulse) cannot stimulate an atomic transition from a z-angular momentum state $j\hbar$ to a state $(j+1)\hbar$, while a σ^+ pulse is the most efficient at this population transfer, regardless of macroscopic phase gradients in the beam. It is thus apparent that the spin (intrinsic) angular momentum is associated with polarization. The interaction strength is proportional to $E_+ = \boldsymbol{E} \cdot \boldsymbol{\sigma}^{+*}$ or $E_- = \boldsymbol{E} \cdot \boldsymbol{\sigma}^{-*}$ at the location of the atom. Thus it is natural to define classical density of spin angular momentum in a uniformly circularly polarized light field as being

$$s_z \propto |E_+|^2 - |E_-|^2 \tag{B.29}$$

For a mode that is entirely right or left polarized, the total spin angular momentum $S_z = \int s_z \, dV$ will then be proportional to the total energy. The proportionality constant turns out to be $1/\omega$ so that the quantized values of S_z are $+\hbar$ and $-\hbar$ according to the ratio argument at the beginning of Sec. B.1. The remaining angular momentum per photon, $(m-s)\hbar = l\hbar$, must be extrinsic, or orbital. One can see that

---

[1]The extrinsic angular momentum is the OAM. Ref. [112] points out that the integrated orbital angular momentum divided by the integrated energy, that is, the OAM per photon, of an apertured beam is "intrinsic" if the beam is apertured by a circle about the beam axis. Taken to the extreme of a circle as small as an atom, this would imply an atom exactly at the axis of the beam could receive angular momentum. However, if this were true, the atom would be able to resolve the axis of the beam to within an Angstrom, which is obviously incorrect. Moreover, orbital angular momentum of light can only be "intrinsic" in exactly the same sense that the angular momentum of a spinning axisymmetric rigid body is intrinsic. (For such a body with angular frequency Ω, the angular momentum and the kinetic energy of constituent particles, integrated over the same cylindrical volume centered on the axis of rotation z, have a constant ratio $\int J_z \, dV / \int U \, dV = 2/\Omega$, regardless of the radius of the cylindrical volume, which can be smaller than the radius of the body, and independent of parallel shifts of the z axis, around which J_z is measured.)

mixed modes do not have a well-defined spin or orbital angular momentum, since S_z is not proportional to the energy.

Discussions of the separation of orbital and spin angular momentum are given in Refs. [22, 108, 109] as well as in Section 10.6 of the book by Mandel and Wolf [113], and in Exercises 7.27-7.29 of Jackson [8]. It is traditional to first introduce the orbital angular momentum of beams for linearly polarized beams such as $(LG_p^l \sigma^+ + LG_p^l \sigma^-)$ or for arbitrarily polarized beams. We will point out one defining feature of all beams with OAM: because of the $e^{il\phi}e^{ikz}$ dependence of certain field components, these field components of the beam have helical phase fronts.

Appendix C

STACK AND MIXING DATA

This appendix gives various data for the mixing calculations for cavities C1, C2, and C3. Cavities C1, C2, and C3, as well as most of the quantities given in the tables, are introduced in Sec. 3.3. The symbol μ after a value indicates that the number should be multiplied by 10^{-6}. The headings for the columns are as follows: k_d is the design wavenumber for the dielectric stack. k_{DC} is the wavenumber splitting $k_D - k_C$. Im k_D is not given since its small difference from Im k_C is not of interest. All of the k quantities have formal units of radians/micron. z_1 is in microns and is calculated via Eq. (4.2) with the appropriate integer q. ϕ_0, ϵ_s and ϵ_p are defined in Eq. (3.44). These are found by a separate stack calculation in which $\phi_s = \arg r_s$ and $\phi_p = \arg r_p$ are calculated using design wavenumber k_d and incident plane wave wavenumber $k = \text{Re}\, k_C$ at three small θ_k values. ϕ_0 is in radians while the ϵ quantities are in inverse radians. Δk_r is the dimensionless $[(k_C + k_D)/2 - k_d]/k_d$ used in Fig. 3.8. The α quantities are given in degrees. α_1 through α_4 are the four mixing angle results whose order is given in Eq. (3.41). α is the mean of α_1-α_4 and is the value used in the many mixing angle plots. The column "$1 - |r|$" gives the value of one minus the normal-incidence field reflectivity of the mirror at design wavelength k_d and incident wavelength k_C. (Note that at normal incidence, $r = r_s = r_p$.) As is stated in Sec. 3.3.2, all mirror characteristics, including $|r|$, ϕ_0, and $\epsilon_{s/p}$ depend on Δk_r but not on k_d and k individually.

k_d	Re k_C	Re k_{DC}	Im k_C	z_1	ϕ_0	ϵ_p	$\epsilon_s - \epsilon_p$
15.840	15.68009	111μ	-1.091μ	-0.185	-0.636	-2.673	-0.423
15.760	15.66720	71.7μ	-1.00μ	-0.195	-0.377	-2.904	-0.263
15.712	15.65920	45.4μ	-0.974μ	-0.204	-0.216	-3.020	-0.151
15.687	15.65498	31.7μ	-0.964μ	-0.198	-0.132	-3.069	-0.095
15.671	15.65227	23.3μ	-0.960μ	-0.203	-0.077	-3.099	-0.053
15.655	15.64954	16.4μ	-0.956μ	-0.204	-0.023	-3.122	-0.016
15.647	15.64825	14.1μ	-0.954μ	-0.121	-0.005	-3.132	-0.003
15.645	15.64783	13.8μ	-0.953μ	-0.204	0.012	-3.135	0.007
15.640	15.64698	13.5μ	-0.952μ	-0.203	0.029	-3.142	0.020
15.637	15.64647	13.4μ	-0.951μ	-0.204	0.039	-3.148	0.030
15.633	15.64578	14.0μ	-0.950μ	-0.204	0.053	-3.151	0.036
15.628	15.64492	15.1μ	-0.949μ	-0.204	0.070	-3.158	0.049
15.618	15.64321	18.9μ	-0.949μ	-0.202	0.104	-3.168	0.072
15.608	15.64149	23.3μ	-0.950μ	-0.199	0.138	-3.178	0.098
15.592	15.63875	31.6μ	-0.953μ	-0.201	0.193	-3.191	0.138
15.570	15.63498	43.9μ	-0.961μ	-0.199	0.268	-3.201	0.190
15.535	15.62901	63.5μ	-0.981μ	-0.197	0.387	-3.204	0.269

k_d	Δk_r	α_1	α_2	α_3	α_4	α	$1 - |r|$
15.840	-0.0101	41.2	40.9	41.0	41.1	41.0	21.4μ
15.760	-0.0059	39.3	39.0	39.1	39.2	39.1	19.8μ
15.712	-0.0034	36.1	35.8	35.8	36.0	35.9	19.3μ
15.687	-0.0020	32.1	31.8	31.9	32.0	32.0	19.1μ
15.671	-0.0012	27.1	26.8	26.9	27.0	26.9	19.0μ
15.655	-349μ	17.1	16.9	16.9	17.1	17.0	19.0μ
15.647	54μ	9.6	9.4	9.4	9.6	9.5	19.0μ
15.645	182μ	6.7	6.6	6.5	6.7	6.6	19.0μ
15.640	446μ	0.7	0.6	0.2	0.6	0.5	19.0μ
15.637	605μ	-3.5	-3.7	-3.7	-3.4	-3.6	19.0μ
15.633	818μ	-8.5	-8.5	-8.6	-8.3	-8.5	19.1μ
15.628	0.0011	-14.0	-14.1	-14.2	-13.8	-14.0	19.1μ
15.618	0.0016	-22.4	-22.3	-22.6	-22.1	-22.3	19.1μ
15.608	0.0021	-27.7	-27.7	-27.9	-27.5	-27.7	19.2μ
15.592	0.0030	-32.8	-32.8	-33.0	-32.6	-32.8	19.3μ
15.570	0.0042	-36.5	-36.5	-36.7	-36.3	-36.5	19.4μ
15.535	0.0061	-39.3	-39.3	-39.5	-39.1	-39.3	19.9μ

Table C.1: C1 data.

k_d	$\operatorname{Re} k_C$	$\operatorname{Re} k_{DC}$	$\operatorname{Im} k_C$	z_1	ϕ_0	ϵ_p	$\epsilon_\mathrm{s} - \epsilon_\mathrm{p}$
16.20	15.81236	53.8μ	-2.33μ	-0.026	2.986	-0.427	0.1736
16.10	15.81027	40.6μ	-1.65μ	-0.025	3.028	-0.377	0.1258
16.00	15.80831	29.0μ	-1.32μ	-0.024	3.067	-0.339	0.0818
15.90	15.80640	19.8μ	-1.17μ	-0.023	3.105	-0.310	0.0397
15.85	15.80545	16.8μ	-1.14μ	-0.023	3.124	-0.299	0.0190
15.805	15.80459	15.5μ	-1.13μ	-0.023	-3.142	-0.289	0.0001
15.80	15.80451	15.6μ	-1.13μ	-0.023	-3.140	-0.289	-0.0017
15.75	15.80356	16.6μ	-1.14μ	-0.023	-3.121	-0.280	-0.0226
15.7	15.80259	19.7μ	-1.18μ	-0.023	-3.101	-0.273	-0.0438
15.65	15.80161	24.2μ	-1.24μ	-0.024	-3.082	-0.268	-0.0654
15.55	15.79956	35.3μ	-1.49μ	-0.025	-3.041	-0.263	-0.1109
15.35	15.79488	64.7μ	-3.43μ	-0.030	-2.947	-0.288	-0.2193

k_d	Δk_r	α_1	α_2	α_3	α_4	α	$1 - |r|$
16.20	-0.0239	-37.0	-37.0	-37.2	-36.8	-37.0	46.1μ
16.10	-0.0180	-34.1	-34.1	-34.3	-33.9	-34.1	32.7μ
16.0	-0.0120	-29.2	-29.2	-29.4	-29.0	-29.2	26.3μ
15.9	-0.0059	-19.6	-19.5	-19.8	-19.3	-19.5	23.3μ
15.85	-0.0028	-11.2	-11.2	-11.4	-11.0	-11.2	22.7μ
15.805	0.7μ	-1.6	-2.0	-1.8	-1.4	-1.7	22.9μ
15.80	286μ	-0.7	-0.6	-0.8	-0.5	-0.6	22.9μ
15.75	0.0034	10.2	10.1	10.0	10.3	10.1	23.2μ
15.70	0.0065	18.7	18.6	18.5	18.8	18.6	23.9μ
15.65	0.0097	24.7	24.5	24.5	24.7	24.6	25.3μ
15.55	0.0161	31.7	31.4	31.5	31.6	31.6	30.6μ
15.35	0.0290	37.8	37.6	37.6	37.7	37.7	70.0μ

Table C.2: C2 data.

k_d	$\mathrm{Re}\,k_C$	$\mathrm{Re}\,k_{DC}$	$\mathrm{Im}\,k_C$	z_1	ϕ_0	ϵ_p	$\epsilon_\mathrm{s} - \epsilon_\mathrm{p}$
7.950	7.83833	249μ	-1.09μ	-0.336	-0.867	-2.433	-0.538
7.870	7.81683	145μ	-0.883μ	-0.386	-0.431	-2.86	-0.295
7.830	7.80546	82μ	-0.838μ	-0.396	-0.202	-3.03	-0.141
7.810	7.79968	49.9μ	-0.826μ	-0.402	-0.085	-3.10	-0.059
7.800	7.79677	35.5μ	-0.822μ	-0.405	-0.027	-3.12	-0.020
7.793	7.79486	28.1μ	-0.818μ	-0.405	0.015	-3.14	0.010
7.788	7.79326	24.3μ	-0.814μ	-0.405	0.044	-3.15	0.030
7.783	7.79194	23.4μ	-0.811μ	-0.406	0.070	-3.16	0.049
7.773	7.78900	31.2μ	-0.806μ	-0.388	0.127	-3.17	0.089
7.760	7.78503	50.1μ	-0.808μ	-0.406	0.208	-3.19	0.144
7.690	7.76463	163μ	-0.884μ	-0.386	0.613	-3.16	0.408

k_d	Δk_r	α_1	α_2	α_3	α_4	α	$1 - |r|$
7.950	-0.0140	41.7	41.1	41.3	41.5	41.4	20.8μ
7.870	-0.0067	39.7	39.2	39.4	39.6	39.5	17.1μ
7.830	-0.0031	36.1	35.6	35.7	35.9	35.8	16.3μ
7.810	-0.0013	30.3	29.9	29.9	30.3	30.1	16.2μ
7.800	-412μ	23.5	23.1	23.1	23.5	23.3	16.2μ
7.793	180μ	15.5	15.7	15.1	16.0	15.6	16.2μ
7.788	678μ	5.7	6.9	5.2	7.2	6.3	16.2μ
7.783	0.0011	-4.2	-7.6	-4.4	-7.2	-5.8	16.3μ
7.773	0.0020	-21.1	-21.2	-21.5	-20.8	-21.2	16.4μ
7.760	0.0032	-31.7	-31.6	-32.1	-31.2	-31.6	16.5μ
7.690	0.0097	-41.5	-41.6	-41.9	-41.3	-41.6	18.4μ

Table C.3: C3 data.

Appendix D

ELECTROMAGNETIC PERTURBATION THEORY FOR CONDUCTING CAVITIES

As mentioned Sec. 3.2.2, three perturbative calculations which combine to give k for the pure LGσ modes of conducting cavities have been developed by Erickson, Cullen, Yu, and Luk [36, 49, 44, 45, 46, 47]. The three perturbations that must be calculated are

1. the perturbation that arises from neglecting $\partial^2 \psi / \partial z^2$ in the paraxial approximation

2. the perturbation that comes from the deviation of the spherical shape of M2 from the actual $\psi = 0$ wave front of the scalar wave

3. the perturbation that comes from the fact that the correct M2 boundary condition is $\boldsymbol{E}_\parallel = 0$ not $\boldsymbol{E}_{\mathrm{T}} = 0$ (which is represented by $\psi = 0$)

The first two of these perturbations can be handled mostly with a scalar field, while the last contribution contains the additional features of the vector wave. We do not go through these perturbative calculations here, as each one is somewhat involved. Instead we develop a general energy-frequency relation for electromagnetic cavity perturbation problems. This relation is explicitly used in the second perturbative calculation listed above, and, in some sense, is used in the third perturbative calculation as well.

D.1 Electromagnetic Perturbation Theory: The Energy-Frequency Relation

The development here partially follows the textbook of Argence and Kahan [114]. Consider a closed conducting cavity with one excited eigenmode of angular frequency ω. If the cavity is adiabatically deformed so that the volume changes by δV, we expect that the final state will be an eigenmode of the new cavity which has some frequency $\omega + \delta\omega$. We also expect that such a deformation will do work

$$\delta W = -\int_{\delta V} P \, dV, \tag{D.1}$$

on the electromagnetic field of the mode (which has stored energy W). Here P is the electromagnetic pressure that the unperturbed mode exerts on the cavity wall at the location of the deformation. Using a single quantum excitation of the field, we have

$$W = \hbar\omega_0,$$
$$W + \delta W = \hbar(\omega_0 + \delta\omega),$$
$$\Rightarrow \quad \frac{\delta\omega}{\omega_0} = \frac{\delta W}{W}. \tag{D.2}$$

Equation (D.2) relates the change in frequency to the change in energy of the mode, and is sometimes called the action theorem because it means that the quantity W/ω, which has units of action, is constant (its total differential is 0). The action theorem is not directly helpful because unless we know P, which we will calculate shortly. (One might guess that, for a localized deformation, δW is just the unperturbed, time-averaged energy density at the location of the deformation, $w_e + w_m$, multiplied by δV. However, one can immediately see that something is wrong with this perturbation model, because it predicts that the lengthening of a Fabry-Perot cavity will increase the energy and the frequency.)

The calculation of the pressure P for a cavity mode is not a simple as the calculation of the radiation pressure of a beam, which is simply $w_e + w_m$. Here we will calculate P indirectly by carefully constructing a perturbation theory for deformations of the cavity wall. Assuming the interior of the cavity is free space, Maxwell's

equations yield

$$\nabla \times \boldsymbol{E} = i\omega\mu_0\boldsymbol{H}, \tag{D.3}$$

$$\nabla \times \boldsymbol{H}^* = i\omega\epsilon_0\boldsymbol{E}^*. \tag{D.4}$$

Multiplying (D.3) by $\boldsymbol{H}^*$ and (D.4) by $\boldsymbol{E}$ and subtracting yields

$$\boldsymbol{H}^* \cdot \nabla \times \boldsymbol{E} - \boldsymbol{E} \cdot \nabla \times \boldsymbol{H}^* = i\omega(\mu_0\boldsymbol{H} \cdot \boldsymbol{H}^* - \epsilon_0\boldsymbol{E} \cdot \boldsymbol{E}^*). \tag{D.5}$$

It is easy to show that

$$\nabla \cdot (\boldsymbol{E} \times \boldsymbol{H}^*) = \boldsymbol{H}^* \cdot \nabla \times \boldsymbol{E} - \boldsymbol{E} \cdot \nabla \times \boldsymbol{H}^*. \tag{D.6}$$

Note that $(1/2)\boldsymbol{E} \times \boldsymbol{H}^* = \boldsymbol{S}^*$ where $\boldsymbol{S}$ is the complex Poynting vector. Integrating over the cavity volume and using Gauss' theorem gives

$$\int \boldsymbol{E} \times \boldsymbol{H}^* \cdot \mathrm{d}\boldsymbol{s} = i\omega \int (\mu_0\boldsymbol{H} \cdot \boldsymbol{H}^* - \epsilon_0\boldsymbol{E} \cdot \boldsymbol{E}^*)\,\mathrm{d}V. \tag{D.7}$$

The surface is taken to be just inside the cavity walls. As no net energy is expected to be entering or leaving the cavity interior, the left hand side of (D.7) is zero. Thus we have, for the original cavity,

$$i\omega \int_V (\mu_0\boldsymbol{H} \cdot \boldsymbol{H}^* - \epsilon_0\boldsymbol{E} \cdot \boldsymbol{E}^*)\,\mathrm{d}V = 0. \tag{D.8}$$

This is a variational principle for conducting cavities, a special case of a more general variational principle [115]. Note that this yields

$$\int_V \mu_0|H|^2\,\mathrm{d}V = \int_V \epsilon_0|E|^2\,\mathrm{d}V, \tag{D.9}$$

which, upon multiplication by $1/4$, is simply

$$W_{\mathrm{m}} = W_{\mathrm{e}}. \tag{D.10}$$

This equality of the time-averaged stored magnetic and electric energies is quite valuable in the study of closed cavity modes. If we consider a cavity that is slightly

deformed, we have, in place of (D.8),

$$i(\omega + \delta\omega) \int_{V+\delta V} [\mu_0(\boldsymbol{H} + \boldsymbol{\delta H}) \cdot (\boldsymbol{H}^* + \boldsymbol{\delta H}^*) - \epsilon_0(\boldsymbol{E} + \boldsymbol{\delta E}) \cdot (\boldsymbol{E}^* + \boldsymbol{\delta E}^*)] \, dV = 0.$$

(D.11)

The prefactor can immediately be divided away. Keeping terms that are zero order and first order in the differential quantities, and using (D.9), yields

$$\int_{\delta V} [\mu_0|H|^2 - \epsilon_0|E|^2] \, dV + 2 \int_V [\mu_0 \text{Re}(\boldsymbol{H}^* \cdot \boldsymbol{\delta H}) - \epsilon_0 \text{Re}(\boldsymbol{E}^* \cdot \boldsymbol{\delta E}) \, dV = 0. \quad \text{(D.12)}$$

Now we use Maxwell's equations to replace, in the last term, $\boldsymbol{H}^*$ with $(i/\omega\mu_0)\boldsymbol{\nabla} \times \boldsymbol{E}^*$ and $\boldsymbol{\delta E}$ with $[(i/\omega\epsilon_0)\boldsymbol{\nabla} \times \boldsymbol{\delta H} - (\delta\omega/\omega)\boldsymbol{E}]$. (The latter replacement comes from conjugating (D.4) and taking its total differential.) Equation (D.12) thus becomes

$$\int_{\delta V} [\mu_0|H|^2 - \epsilon_0|E|^2] \, dV + 2\text{Re} \int_V \left(\frac{i}{\omega}\right) \left(\boldsymbol{\delta H} \cdot (\boldsymbol{\nabla} \times \boldsymbol{E}^*) - \boldsymbol{E}^* \cdot (\boldsymbol{\nabla} \times \boldsymbol{\delta H})\right) \, dV$$
$$+ 2\text{Re} \int_V \epsilon_0 \frac{\delta\omega}{\omega} |E|^2 \, dV = 0. \quad \text{(D.13)}$$

Using $\boldsymbol{\nabla} \cdot (\boldsymbol{E}^* \times \boldsymbol{\delta H}) = \boldsymbol{\delta H} \cdot (\boldsymbol{\nabla} \times \boldsymbol{E}^*) - \boldsymbol{E}^* \cdot (\boldsymbol{\nabla} \times \boldsymbol{\delta H})$, the second term becomes

$$-\frac{2}{\omega} \text{Im} \int_V \boldsymbol{\nabla} \cdot (\boldsymbol{E}^* \times \boldsymbol{\delta H}) \, dV. \quad \text{(D.14)}$$

Using Gauss' theorem the integral becomes

$$\int_S (\boldsymbol{E}^* \times \boldsymbol{\delta H}) \cdot \mathbf{ds}. \quad \text{(D.15)}$$

This integral must be zero because the tangential electric field just off the surface of a conductor is zero, and hence $\boldsymbol{E} \parallel \mathbf{ds}$. Noting that the integrand in the last term of (D.13) is already real, the equation becomes

$$\int_{\delta V} [\mu_0|H|^2 - \epsilon_0|E|^2] \, dV + 2\frac{\delta\omega}{\omega} \int_V \epsilon_0|E|^2 \, dV. \quad \text{(D.16)}$$

Using (D.9), we can write the energy-frequency relation in a final form

$$\frac{\delta\omega}{\omega} = -\frac{\int_{\delta V}\left(\frac{1}{4}\mu_0|H|^2 - \frac{1}{4}\epsilon_0|E|^2\right)\mathrm{d}V}{\int_V\left(\frac{1}{4}\mu_0|H|^2 + \frac{1}{4}\epsilon_0|E|^2\right)\mathrm{d}V}. \tag{D.17}$$

The factors of $1/4$ have been included so that the denominator is W, the time-averaged stored energy. From (D.1) we see that the local pressure is

$$P = w_{\mathrm{m}} - w_{\mathrm{e}}, \tag{D.18}$$

the difference of the time-averaged magnetic and electric energy densities.

Finally we note that, for paraxial modes in a Fabry-Perot cavity, the electric energy density near the mirror, $w_{\mathrm{e}} = (1/4)\epsilon_0|E_\perp|^2$, is quite small (it is order h, since E_z is order h). This means that w_{e} can be neglected and only $w_{\mathrm{m}} \approx (1/4)\mu_0\boldsymbol{H}_{\mathrm{T}} \cdot \boldsymbol{H}_{\mathrm{T}}^*$ need be used in the calculation.

Appendix E

SOPHISTICATED CALCULATION OF THE GH EFFECT AT A DIELECTRIC INTERFACE

We follow and expand on the careful derivations of Lai, Cheng, and Tang [85], who expand on work by Horowitz and Tamir [116]. We only consider s polarization and reprint Eq. (5.19) here as

$$
\psi_{\text{refl}}(x_2, z_2) = \int_{-\infty}^{\infty} \tilde{\psi}(k_x) \underbrace{r_s[\theta_k(k_x)]e^{i\sqrt{k_A^2 - k_x^2}z_2}}_{S(k_x)} e^{ik_x x_2} \, dk_x. \tag{E.1}
$$

Note the definition of $S(k_x)$. The first approximation has been to take the integration boundaries to ∞. The goal is to evaluate (approximate) the integral, when the incident beam is Hermite-Gaussian, in a way which is valid both near and far away from the critical angle. The Taylor series expansion of ϕ_s used in Eq. (5.21) is singular at the critical angle, and thus a modified procedure is necessary. Instead of working with $\arg[S(k_x)]$ we will work with $S(k_x)$ itself, expanding it in terms of k_x/k_A. This Taylor expansion also contains singular terms at the critical angle, so these terms are dealt with in a special way, resulting in the form $(-\Delta - k_x/k_A)^{1/2}$, where Δ is related

to χ and χ_c by

$$\Delta \equiv \frac{\sin^2 \chi - \sin^2 \chi_c}{\sin 2\chi},$$ (E.2)

$$\Delta \approx \chi - \chi_c \quad \text{when } \chi - \chi_c \ll 1.$$ (E.3)

The notation for the expansion of $S(k_x)$ will be

$$S(k_x) = \Big[\sum_{n=0}^{n_S} \Big(\frac{k_x}{k_A}\Big)^n S_n\Big] + \Big[\sum_{n=0}^{n_T} \Big(\frac{k_x}{k_A}\Big)^n \Big(-\Delta - \frac{k_x}{k_A}\Big)^{1/2} T_n\Big].$$ (E.4)

In contrast to Artmann's formulation, it will not be possible to determine the GH shift before performing the k_x integral. The integral will be evaluated for each type of term in the expansion (E.4). The result will be

$$\psi_{\text{refl}} = \tilde{A}_0\Big[\sum_{n=0}^{n_S} S_n s_n(x_2)\Big] + \tilde{A}_0\Big[\sum_{n=0}^{n_T} T_n t_n(x_2)\Big],$$ (E.5)

where s_n and t_n are integrals which will be precisely defined later and $\tilde{A}_0$ is a constant, also defined later. After evaluating ψ_{refl}, the GH shift will be determined by the shift of the maximum of $|\psi_{\text{refl}}|^2$ away from $x_2 = 0$. We will perform the general version of this last step first, to complete the overall picture of the process.

The function ψ_{refl} will be factorable as

$$\psi_{\text{refl}}(x_2, z_2) = \frac{(-1)^m A_0}{\sqrt{\sigma}\cos^2 \chi_c} e^{ik_A z_2} e^{-x_2^2/(2\sigma^2)} f\Big(\frac{x_2}{\sigma}; k_A, \sigma, \chi, \chi_c\Big),$$ (E.6)

where f is to be determined, A_0 is a constant to be defined, m is the order of the Hermite-Gauss beam, and we have introduced the notation

$$\sigma \equiv w_0/\sqrt{2},$$ (E.7)

where w_0 is the waist radius of the beam. Throughout this section we will be taking $(k_A\sigma)^{-1}$, which is proportional to our paraxiality parameter h, to be a small quantity:

$$(k_A\sigma)^{-1} \ll 1.$$ (E.8)

We pay significant attention to approximation errors in our treatment, partly because

in the applications we wish to study $(k_A\sigma)^{-1}$ can be as large as 0.2.

Writing the last two factors of (E.6) as

$$\exp[-x_2^2/(2\sigma^2)]f(x_2) = \exp[-x_2^2/(2\sigma^2) + \ln f], \qquad (E.9)$$

we expand $\ln f$ around $x_2/\sigma = 0$ as

$$\ln f = a_0 + a_1\frac{x_2}{\sigma} + \frac{a_2}{2}\frac{x_2^2}{\sigma^2} + \frac{a_3}{6}\frac{x_2^3}{\sigma^3} + \cdots. \qquad (E.10)$$

This expansion is substituted into the RHS of (E.9). The maximum of $|\psi_{\mathrm{refl}}|^2$ occurs at the value of x_2 where

$$\frac{\mathrm{d}}{\mathrm{d}x_2}\mathrm{Re}\left[-\frac{x_2^2}{2\sigma^2} + \ln f\right] = 0. \qquad (E.11)$$

If only the linear term in (E.10) is kept, the maximum of $|\psi_{\mathrm{refl}}|^2$ is shifted from $x_2 = 0$ to an x_2 value of

$$\Delta_{\mathrm{GH}}^{(L1)} = \sigma\mathrm{Re}\,a_1. \qquad (E.12)$$

Keeping up to the quadratic and cubic terms yields

$$\Delta_{\mathrm{GH}}^{(L2)} = \sigma\frac{\mathrm{Re}\,a_1}{1 - \mathrm{Re}\,a_2}, \qquad (E.13)$$

$$\Delta_{\mathrm{GH}}^{(L3)} = \sigma\frac{1 - \mathrm{Re}\,a_2 - \sqrt{(1 - \mathrm{Re}\,a_2)^2 - 2\mathrm{Re}\,a_1\mathrm{Re}\,a_3}}{\mathrm{Re}\,a_1}, \qquad (E.14)$$

respectively. The result used by Lai *et al.* is essentially $\Delta_{\mathrm{GH}}^{(L2)}$. When $(k_A\sigma)^{-1} \ll \Delta$ one can reasonably justify the truncation of (E.10) at quadratic order, as will be shown later. When this is not the case (when $\chi \approx \chi_c$), the best one can do is to calculate the first term being dropped and verify that it is a small correction. In terms of f, the first four a_j coefficients are

$$a_0 = \ln f,$$

$$a_1 = \frac{f'}{f},$$

$$a_2 = -\frac{f'^2}{f^2} + \frac{f''}{f},$$

$$a_3 = 2\frac{f'^3}{f^3} - 2\frac{f'f''}{f^2} - \frac{f'f''}{f^2} + \frac{f'''}{f}, \tag{E.15}$$

where the primes denote derivatives with respect to (x_2/σ) keeping σ constant, and the final expressions are evaluated at $x_2 = 0$.

E.0.1 Integration of Individual Terms

Before performing the (modified) Taylor series expansion of $S(k_x)$, it is helpful to look at how terms of different order will behave when integrated. In the future, we may wish to extend the calculations of Lai *et al.* to a Hermite-Gauss (HG) beam, so we include HG notation here.

The x-dependent part of mode pattern, at the waist, of the HG beam of Eq. (3.8) is

$$\psi(x) = \frac{A_0}{\sqrt{\sigma}} H_m\left(\frac{x}{\sigma}\right) e^{-\frac{x^2}{2\sigma^2}}. \tag{E.16}$$

Of course the fundamental Gaussian has Hermite index $m = 0$. One may consider A_0 to be arbitrary or set it to $\pi^{-1/4}(m!2^m)^{-1/2}$ for normalization. The Fourier transform (using the standard $1/(2\pi)$ prefactor) of ψ is

$$\tilde{\psi}(k_x) = \underbrace{(i^m(2\pi)^{-1/2})A_0}_{\tilde{A}_0} \sqrt{\sigma} H_m(-\sigma k_x) e^{-\sigma^2 k_x^2/2}. \tag{E.17}$$

The transform has been evaluated using equation 7.376.1 of Ref. [48],

$$\int_{-\infty}^{\infty} e^{-x^2/2} H_m(x) e^{iyx}\, dx = \sqrt{2\pi} i^m H_m(y) e^{-y^2/2}. \tag{E.18}$$

(The "Laguerre analog" of this integral is Eq. (3.26).)

The term-by-term evaluation of ψ_{refl} (see Eqs. (E.1, E.4, E.5, E.17)) contains the integrals (inverse FTs)

$$\begin{aligned}
s_n(x_2; \sigma, k_A) &\equiv \int_{-\infty}^{\infty} \frac{\tilde{\psi}(k_x)}{\tilde{A}_0}\left(\frac{k_x}{k_A}\right)^n e^{ix_2 k_x}\, dk_x \\
&= \sqrt{\sigma} \int_{-\infty}^{\infty} \left(\frac{k_x}{k_A}\right)^n H_m(-\sigma k_x) e^{-\sigma^2 k_x^2/2} e^{ix_2 k_x}\, dk_x,
\end{aligned} \tag{E.19}$$

where n is some small power. For $n = 0$, using (E.18) yields

$$s_0(x_2) = i^m \sqrt{\frac{2\pi}{\sigma}} H_m\left(-\frac{x_2}{\sigma}\right) e^{-x_2^2/(2\sigma^2)}. \tag{E.20}$$

To obtain s_n from s_0, the operator

$$\hat{S}_+ \equiv \frac{-i}{k_A} \frac{\mathrm{d}}{\mathrm{d}x_2} = \frac{-i}{k_A \sigma} \left(\frac{\partial}{\partial(x_2/\sigma)}\right)_\sigma \tag{E.21}$$

can be applied n times to s_0. One can see that $k_A \sigma \gg 1 \Rightarrow |s_j| \ll |s_{j-1}|$ by the following facts. The action of $\hat{S}_+$ on $\exp[-x^2/(2\sigma^2)]$ is $(ix_2/\sigma)(k_A\sigma)^{-1} \exp[-x^2/(2\sigma^2)]$ and the action of $\hat{S}_+$ on $(ix_2/\sigma)^l$ is $l(k_A\sigma)^{-1}(ix_2/\sigma)^{l-1}$. The action of $\hat{S}_+$ on $H_m(-x_2/\sigma)$ is $i(k_A\sigma)^{-1}(2m)H_{m-1}$ or $i(k_A\sigma)^{-1}[2(x_2/\sigma)H_m - H_{m+1}]$. Thus the magnitude of s_n scales as $(k_A\sigma)^{-n} \propto h^n$, meaning that one can generally truncate the series $\sum_n S_n s_n(x_2)$ at low n. The Hermite order m does not affect the scaling of s_n. For future reference

$$s_1(x_2) = i^{m+1} \sqrt{\frac{2\pi}{\sigma}} \frac{1}{k_A\sigma} \left[\frac{x_2}{\sigma} H_m\left(-\frac{x_2}{\sigma}\right) + 2m H_{m-1}\left(-\frac{x_2}{\sigma}\right)\right] e^{-x_2^2/(2\sigma^2)}. \tag{E.22}$$

In addition to s_n, the other inverse FT that will arise is

$$t_n(x_2; \sigma, k_A) = \sqrt{\sigma} \int_{-\infty}^{\infty} \left(\frac{k_x}{k_A}\right)^n \left(-\Delta - \frac{k_x}{k_A}\right)^{1/2} H_m(-k_x\sigma) e^{-\sigma^2 k_x^2/2} e^{ix_2 k_x} \, \mathrm{d}k_x. \tag{E.23}$$

With $m = n = 0$, following the technique of Ref. [116] yields

$$t_0(x_2) = \left(\frac{2\pi i}{k\sigma^2}\right)^{1/2} e^{-x_2^2/(2\sigma^2)} e^{\beta^2/4} D_{1/2}(\beta), \tag{E.24}$$

where

$$\beta = i(k_A\sigma)\Delta - \frac{x_2}{\sigma}, \tag{E.25}$$

and $D_p(z)$ is the parabolic cylinder function (see Appendix G). Note that the imaginary part of β contains the large quantity $k_A\sigma$. Exactly as for the s_n quantities, $\hat{S}_+$ acts as raising operator:

$$t_n = (\hat{S}_+)^n t_0. \tag{E.26}$$

For $m = 0$, $n = 1$

$$t_1(x_2) = \frac{i}{k_A\sigma}\frac{x_2}{\sigma}t_0 + \frac{i}{2k_A\sigma}\frac{D_{-1/2}(\beta)}{D_{1/2}(\beta)}t_0, \qquad (E.27)$$

where the recursion relation (G.6) has been used. One can see that $k_A\sigma \gg 1 \Rightarrow |t_j| \ll |t_{j-1}|$ by noting the previous discussion of $\hat{S}_+$ on Gaussian and power terms, as well as noting that the action of $\hat{S}_+$ on $\exp(\beta^2/4)D_\nu(\beta)$ is $i\nu(k_A\sigma)^{-1}\exp(\beta^2/4)D_{\nu-1}(\beta)$. Since the leading term of the asymptotic expansion ($|z| \gg 1, |z| \gg |\nu|$) of $D_\nu(z)$ is $\exp(-z^2/4)z^\nu$ (see (G.3)), this action of $\hat{S}_+$ "brings down" at least one and possibly two powers of $(k_A\sigma)^{-1}$, depending on whether Δ is "small".

Thus it has been successfully shown that factors of $(k_x/k_A)^n$ in integrands of Eqs. (E.23) and (E.19) cause the integrals to be of order $(k_A\sigma)^{-n}$, justifying a number of approximations. In all cases it is known how to calculate the exact size of the largest term that is being dropped[1], a calculation which may be necessary when $k_A\sigma$ is not far greater than 1. Since the argument of the Hermite polynomials is $(-k_x/k_A) \times (k_A\sigma)$, the Hermite order m does not contribute to the smallness of the integrals.

E.0.2 Approximating $S(k_x)$

The next step is to approximately expand $S(k_x)$, the active ingredient of (E.1), in k_x/k_A. As mentioned previously, some of the terms will have to be treated specially due to singularities.

To second order in k_x/k_A, the expansion of the z_2-dependent exponential factor is

$$\exp\left(iz_2\sqrt{k_A^2 - k_x^2}\right) \approx \left[1 - i\frac{k_A z_2}{2}\left(\frac{k_x}{k_A}\right)^2\right]e^{ik_A z_2}. \qquad (E.28)$$

From the general rule that $(k_x/k_A)^n \rightarrow (k_A\sigma)^{-n}$ upon integration, we see that the second term in the brackets can be dropped[2] if

$$z_2 \ll 2(k_A\sigma)\sigma. \qquad (E.29)$$

If needed, the coefficient (presumably of order one) of the right hand side can be

---

[1]This calculation should not increase computation time a great deal. The parabolic cylinder functions $D_{-3/2}, D_{-5/2}, \ldots$ can be quickly obtained from $D_{-1/2}$ and $D_{1/2}$ via a recursion relation.

[2]Lai et al. give the more favorable condition $z_2^2 \ll (k_A\sigma)^2\sigma^2$. An examination of our final results indicates that this weaker condition appears to be sufficient if $\Delta k_A\sigma \gg 1$.

calculated. The value of z_2 is however somewhat vague. As we wish to apply our calculations to a beam bouncing within a cavity, the z_2 values roughly correspond to the half-lengths of individual trajectory segments. (If the Gaussian ray model is to be reasonable, we must imagine the Gaussian beam as being refocused at each bounce so that each trajectory segment contains a waist region. Given that the focal length of a spherical (or cylindrical) mirror is $R_c/2$, this conjecture of focusing behavior is loosely supported: a six bounce periodic orbit (hexagon) in a circle has segments of length R_c, so that the nominal beam focus of $z_2 = R_c/2$ is at the center of each segment. We have not included any focusing effects, such as using beam radius $w(z_2)$ instead of w_0, in our analysis.) The condition (E.29) is somewhat in conflict with a condition of our sequential tunneling model, that the width of the beam be small compared to the resonator so that a small portion of the boundary is subtended by the beam upon reflection.

If, on the other hand, one does not drop the quadratic term in (E.28), the predicted GH shift will depend on z_2. This shift will, in the far field, become linear in z_2 and will be described not as a shift, but as a change in the angle of the reflected beam. Namely, it is the Fresnel kick. We calculate the Fresnel kick (in the far field) separately in the next section. Unfortunately, a full calculation for intermediate z_2 values appears to be difficult. We drop the quadratic term in the GH calculation, and in doing so limit the expansion of r_s to order $(k_x/k_A)^1$ or order $(k_x/k_A)^{3/2}$, depending on whether Δ is small.

Finally we expand r_s, the other factor in $S(k_x)$. Rationalizing r_s in (5.23) by multiplying numerator and denominator by $\cos\theta_k - [\sin^2\chi_c - \sin^2\theta_k]^{1/2}$ yields

$$r_s = \frac{\cos^2\theta_k - \sin^2\theta_k + \sin^2\chi_c - 2\cos\theta_k[\sin^2\chi_c - \sin^2\theta_k]^{1/2}}{\cos^2\chi_c}. \tag{E.30}$$

Note that

$$\theta_k = \chi + \arcsin\frac{k_x}{k_A},$$

$$\cos\theta_k = \sqrt{1 - \frac{k_x^2}{k_A^2}}\cos\chi - \frac{k_x}{k_A}\sin\chi,$$

$$\sin\theta_k = \sqrt{1 - \frac{k_x^2}{k_A^2}}\sin\chi + \frac{k_x}{k_A}\cos\chi. \tag{E.31}$$

Inserting these into the numerator of (E.30), N_r, and using a number of trigonometric identities, yields without approximation:

$$N_r = N_{r1} + N_{r2},$$

$$N_{r1} = \cos^2\chi - \Delta\sin 2\chi - 2\frac{k_x}{k_A}\sqrt{1 - \frac{k_x^2}{k_A^2}}\sin 2\chi - 2\frac{k_x^2}{k_A^2}\cos 2\chi,$$

$$N_{r2} = -2\sqrt{\sin 2\chi}\left(\sqrt{1 - \frac{k_x^2}{k_A^2}}\cos\chi - \frac{k_x}{k_A}\sin\chi\right)\left(-\Delta - \frac{k_x}{k_A}\sqrt{1 - \frac{k_x^2}{k_A^2}} - \frac{k_x^2}{k_A^2}\cot 2\chi\right)^{1/2}.$$

$$\text{(E.32)}$$

Immediately dropping terms of order greater than two in each factor yields

$$N_{r1} \approx \cos^2\chi - \Delta\sin 2\chi - 2\frac{k_x}{k_A}\sin 2\chi - 2\frac{k_x^2}{k_A^2}\cos 2\chi,$$

$$N_{r2} \approx -2\sqrt{\sin 2\chi}\left[(1 - \frac{k_x^2}{2k_A^2})\cos\chi - \frac{k_x}{k_A}\sin\chi\right]\left(-\Delta - \frac{k_x}{k_A} - \frac{k_x^2}{k_A^2}\cot 2\chi\right)^{1/2}. \quad \text{(E.33)}$$

At this point it is helpful to know the ranges of the trigonometric coefficients. The possible range of χ is $[0, \pi/2)$ and we are not interested in χ very close to either end of this range. This means that $\sin\chi$, $\cos\chi$, and $\sin 2\chi$ never go to 0, while $\cos 2\chi$ is zero at $\chi = \pi/4$. The constant term in N_r,

$$f_0 \equiv \cos^2\chi - \Delta\sin 2\chi = \cos 2\chi + \sin^2\chi_c, \quad \text{(E.34)}$$

can be zero for $\chi > \pi/4$ and sufficiently large $\chi - \chi_c$. For this reason we keep at least one term in addition to f_0 in every approximation to N_r. Dropping the last term of N_{r1} is justified if

$$(k_A\sigma)^{-1} \ll |\tan 2\chi|, \quad \text{(E.35)}$$

even if f_0 is 0. The numerical prefactor of the RHS in this case could be calculated with the integral relations previously given. Assuming for the moment that the $(k_x/k_A)^2$ term inside the square root factor of N_{r2} has been dropped, we see that dropping other $(k_x/k_A)^2$ term in N_{r2} relies on condition (E.35), where again a (different) numerical prefactor could be calculated for the RHS. It is reasonable that the approximation of dropping the $(k_x/k_A)^2$ term inside the square root factor also depends on condition

(E.35). Here we are assuming that the $k_x/k_A \to (k_A\sigma)^{-1}$ replacement is valid within the square root factor and the prefactor involved in the replacement is of order one; in this case we cannot calculate this prefactor, or the prefactor of the RHS of (E.35), as we do not have an analytic expression for the inverse FT containing the entire last factor of N_{r2}. Thus N_r becomes

$$
N_r \approx \left[\cos^2\chi - \Delta\sin 2\chi\right] + \left[-2\sin 2\chi\right]\frac{k_x}{k_A}
$$
$$
+ \left[-2\sqrt{\sin 2\chi}\cos\chi\right]\left(-\Delta - \frac{k_x}{k_A}\right)^{1/2} + \left[2\sqrt{\sin 2\chi}\sin\chi\right]\frac{k_x}{k_A}\left(-\Delta - \frac{k_x}{k_A}\right)^{1/2}.
$$
$$
\tag{E.36}
$$

The four expressions in square brackets, when multiplied by $e^{ik_Az_2}/\cos^2\chi_c$, correspond respectively to S_0, S_1, T_0, and T_1 of expansion (E.4). When $\Delta = 0$, the last term of (E.36) is of order $(k_x/k_A)^{3/2}$, and we have already dropped another term of this order. It could thus be argued that the factor $(-\Delta - k_x/k_A)^{1/2}$ in the last term should be replaced with $(-\Delta)^{1/2}$. However, we will follow Lai *et al.* and keep this term to prevent the final result from having an infinite derivative at $\Delta = 0$.

E.0.3 The GH Shift for the Fundamental Gaussian

For a fundamental Gaussian ($m = 0$), combining previous expressions results in the function $f(x_2/\sigma)$ (defined in Eq. (E.6)) being given by

$$
f = f_0 - 2i^{1/2}\cos\chi\sqrt{\sin 2\chi}(k_A\sigma)^{-1/2}e^{\beta^2/4}D_{1/2}(\beta) - 2i\sin(2\chi)(k_A\sigma)^{-1}\frac{x_2}{\sigma}
$$
$$
+ 2i^{3/2}\sin\chi\sqrt{\sin 2\chi}(k_A\sigma)^{-3/2}e^{\beta^2/4}\left(\frac{x_2}{\sigma}D_{1/2}(\beta) + \frac{1}{2}D_{-1/2}(\beta)\right). \tag{E.37}
$$

Taking the derivative with respect to x_2/σ holding σ constant yields

$$
f' = i^{1/2}\cos\chi\sqrt{\sin 2\chi}(k_A\sigma)^{-1/2}e^{\beta^2/4}D_{-1/2}(\beta) - 2i\sin(2\chi)(k_A\sigma)^{-1}
$$
$$
+ 2i^{3/2}\sin\chi\sqrt{\sin 2\chi}(k_A\sigma)^{-3/2}e^{\beta^2/4}\left(\frac{3}{2}D_{1/2}(\beta) - \frac{i\Delta}{2}k_A\sigma D_{-1/2}(\beta)\right). \tag{E.38}
$$

The second derivative yields

$$f'' = \frac{i^{1/2}}{2}\left(\cos\chi + \Delta\sin\chi\right)\sqrt{\sin 2\chi}(k_A\sigma)^{-1/2}e^{\beta^2/4}D_{-3/2}(\beta)$$
$$- \frac{3i^{3/2}}{2}\sin\chi\sqrt{\sin 2\chi}(k_A\sigma)^{-3/2}e^{\beta^2/4}D_{-1/2}(\beta). \tag{E.39}$$

In further derivatives these two terms proportional to parabolic cylinder functions persist, with each derivative of $\exp(\beta^2/4)D_{-l/2}(\beta)$ being $(l/2)\exp(\beta^2/4)D_{-l/2-1}(\beta)$. To calculate the GH shift, f, f', f'', and perhaps higher derivatives are evaluated at $x_2 = 0$ (so that $\beta = i\Delta k_A\sigma$), and are substituted into (E.15).

When

$$(k_A\sigma)^{-1} \ll \Delta, \tag{E.40}$$

one can see from the asymptotic behavior of $\exp(\beta^2/4)D_p(\beta)$ given in (G.3) that action of each derivative makes the term smaller by $\sim (\Delta k_A\sigma)^{-1}$. In this case $|f^{(j+1)}| \ll |f^{(j)}|$ which leads to $|a_{j+1}| \ll |a_j|$, and the truncation of expansion (E.10) is reasonably justified. On the other hand, when $\Delta = 0$, or is sufficiently small so that $(k_A\sigma)^{-1} \sim \Delta$, the truncation cannot be justified a priori.

Lai et al. drop several terms in the calculation of a_1 and a_2 from f, f', and f'''. Their final expressions are

$$a_1^{(L)} = \left[i^{1/2}\cos\chi\sqrt{\sin 2\chi}(k_A\sigma)^{-1/2}e^{\beta_0^2/4}D_{-1/2}(\beta_0) - 2i\sin(2\chi)(k_A\sigma)^{-1}\right.$$
$$\left. + 2i^{3/2}\sin\chi\sqrt{\sin 2\chi}(k_A\sigma)^{-3/2}e^{\beta_0^2/4}D_{1/2}(\beta_0)\right]$$
$$\times \left[f_0 - 2i^{1/2}\cos\chi\sqrt{\sin 2\chi}(k_A\sigma)^{-1/2}e^{\beta_0^2/4}D_{1/2}(\beta_0)\right]^{-1},$$
$$a_2^{(L)} = \frac{i^{1/2}}{2f_0}\cos\chi\sqrt{\sin 2\chi}(k_A\sigma)^{-1/2}e^{\beta_0^2/4}D_{-3/2}(\beta_0), \tag{E.41}$$

where $\beta_0 = i\Delta k_A\sigma$. It appears that Lai et al. have approximated the expression in large parentheses in (E.38) as $D_{1/2}(\beta)$. This approximation works well if condition (E.40) is met, but obviously causes 33% relative error in the last term of (E.38) when $\Delta = 0$. As it does not add significant computational cost, we use the exact formulae of (E.15, E.37, E.38, E.39) to calculate $\Delta_{GH}^{(L2)}$.

For $\Delta > 0$, Lai et al. show that in the limit $k_A\sigma \to \infty$ the GH shift calculated by the methods given here gives Artmann's result, $\Delta_{GHs}^{(A)}$.

Appendix F

ANGULAR MOMENTUM OF CYLINDER QUASIMODES

Here we discuss the angular momentum and energy of modes in infinite dielectric cylinders of oval cross section. The concepts of angular momentum and energy for a quasimode (k complex) are perhaps not well defined. Nevertheless, we use the standard definitions of energy and momentum densities and discuss the results. As in Appendix B, we wish to show that a quasimode with circular symmetry has angular momentum $m\hbar$ per quanta by showing that

$$\frac{\text{total classical angular momentum}}{\text{total classical energy}} = \frac{m}{\omega}. \qquad (F.1)$$

F.1 Total Angular Momentum

We partition space into regions 1 and 2, with region 1 being inside the dielectric. We take $n_1 = n$ and $n_2 = 1$ and use the notation

$$
\begin{aligned}
Z_m^{(1)}(x) &\equiv J_m(x), \\
Z_m^{(2)}(x) &\equiv H_m^{(1)}(x), \\
a_m^{(1)} &\equiv a_m, \\
a_m^{(2)} &\equiv b_m.
\end{aligned}
\qquad (F.2)
$$

where the a_m and b_m coefficients are the same as in Sec. 1.2.1. The electric field in region j is given by

$$E^{(j)} = \hat{z} \sum_{m=0}^{\infty} a_m^{(j)} Z_m^{(j)}(n_j k\rho) e^{im\phi} e^{-i\omega t}. \tag{F.3}$$

Via the Maxwell equation

$$H = \frac{-i}{\mu\omega} \nabla \times E, \tag{F.4}$$

we obtain

$$H^{(j)} = \frac{1}{\mu_0 \omega} \sum_{m=0}^{\infty} a_m^{(j)} \left[\frac{m}{\rho} G_m^{(j)}(n_j k\rho) \hat{\rho} + i n_j k G_m^{(j)'}(n_j k\rho) \hat{\phi} \right] e^{-i\omega t}. \tag{F.5}$$

The classical time-averaged angular momentum density is

$$m = \frac{n^2}{2c^2} \text{Re}[r \times (E \times H^*)]. \tag{F.6}$$

The z component is

$$m_z = \frac{n^2}{2c^2} \text{Re}(E_z H_\rho^*)\rho, \tag{F.7}$$

$$= \frac{n_j^2 \epsilon_0}{2\omega} \sum_m \sum_{m'} m' \text{Re}[a_m^{(j)} a_{m'}^{(j)*} Z_m^{(j)}(n_j k\rho) Z_{m'}^{(j)*}(n_j k\rho) e^{i(m-m')\phi}] \tag{F.8}$$

The total angular momentum in the infinite volume between $z = 0$ and $z = d > 0$ is

$$M_z = \frac{d\epsilon_0}{2\omega} \sum_{j=1}^{2} \left[n_j^2 \int_{A_j} dA \sum_m \sum_{m'} m' \text{Re}[a_m^{(j)} a_{m'}^{(j)*} Z_m^{(j)}(n_j k\rho) Z_{m'}^{(j)*}(n_j k\rho) e^{i(m-m')\phi}] \right]. \tag{F.9}$$

Here we give some special cases. For a cylinder bounded by a conductor, so that k is real and the mode is entirely inside region 1, we have

$$M_{z,\text{cond}} = \frac{d n_1^2 \epsilon_0}{2\omega} \int_{A_1} dA \sum_m \sum_{m'} m' J_m(n_1 k\rho) J_{m'}(n_1 k\rho) \text{Re}[a_m^{(1)} a_{m'}^{(1)*} e^{i(m-m')\phi}]. \tag{F.10}$$

On the other hand, if the resonator is dielectric (with no conducting boundary),

but is circular, then the solution possesses a single m component, yielding

$$M_{z,\text{circ}} = \frac{md\epsilon_0}{2\omega} \sum_{j=1}^{2} \left[n_j^2 \int_{A_j} dA |a_m^{(j)}|^2 |Z_m^{(j)}(n_j k\rho)|^2 \right].$$
(F.11)

F.2 Total Energy

The total energy is

$$U = U_\text{E} + U_\text{H},$$
(F.12)

where

$$U_{\text{E/H}} = d \sum_{j=1}^{2} \int_{A_j} u_{\text{E/H}}^{(j)} \, dA.$$
(F.13)

We will argue that U_H should be (almost) equal to U_E.

F.2.1 Electric Energy

The time-averaged electric energy density is

$$u_\text{E} \simeq \frac{n^2 \epsilon_0}{4} \boldsymbol{E} \cdot \boldsymbol{E}^*.$$
(F.14)

Substituting (F.3) into this equation yields

$$u_\text{E}^{(j)} = \frac{n_j^2 \epsilon_0}{4} \sum_m \sum_{m'} a_m^{(j)} a_{m'}^{(j)*} Z_m^{(j)}(n_j k\rho) Z_{m'}^{(j)*}(n_j k\rho) e^{i(m-m')\phi}.$$
(F.15)

For the case of cylindrical symmetry

$$u_{\text{E,circ}}^{(j)} = \frac{n_j^2 \epsilon_0}{4} |a_m^{(j)}|^2 |Z_m^{(j)}(n_j k\rho)|^2.$$
(F.16)

Immediately we have the nice property $m_{z,\text{circ}}/u_{\text{E,circ}}^{(j)} = m/(2\omega)$. If it were the case that $u_\text{H} = u_\text{E}$, then we could immediately show Eq. F.1. Unfortunately this is not the case.

F.2.2 Magnetic Energy

The time-averaged magnetic energy density is

$$u_{\mathrm{H}} = \frac{\mu_0}{4} \boldsymbol{H} \cdot \boldsymbol{H}^*. \tag{F.17}$$

This yields

$$u_{\mathrm{H}}^{(j)} = \frac{1}{4\mu_0\omega^2} \sum_m \sum_{m'} a_m^{(j)} a_{m'}^{(j)*} e^{i(m-m')\phi}$$
$$\times \left[\frac{mm'}{\rho^2} G_m^{(j)}(n_j k\rho) G_{m'}^{(j)*}(n_j k\rho) + n_j^2 |k|^2 G_m^{(j)\prime}(n_j k\rho) G_{m'}^{(j)\prime*}(n_j k\rho) \right]. \tag{F.18}$$

When there is cylindrical symmetry, this becomes

$$u_{\mathrm{H}}^{(j)} = \frac{n^2 |k|^2}{8\mu_0\omega^2} |a_m^{(j)}|^2 \left[|G_{m-1}^{(j)}(n_j k\rho)|^2 + |G_{m+1}^{(j)}(n_j k\rho)|^2 \right]. \tag{F.19}$$

We do not know how to show that the integral of u_{H} is equal to U_{E}. However, it appears that $U_H = U_{\mathrm{E}}$ is a general property of electromagnetic resonators driven at a resonance frequency. Indeed the equality can be shown (at resonance frequency) for resonators for which the equivalent circuit is an LRC circuit with a single input/output terminal [117]. The author does not know whether the dielectric cylinder falls into this class. For conducting, closed resonators, Poynting's theorem gives the energy equality directly [114], since the integral version of Poynting's theorem is

$$\oint (\boldsymbol{E} \times \boldsymbol{H}^*) \cdot \mathrm{d}\boldsymbol{S} = 4i\omega(U_{\mathrm{E}} - U_{\mathrm{H}}). \tag{F.20}$$

Taking the surface to be the conductor, the integrand on the left is always zero. Even if one does not have a conducting surface, we can construct a surface around the outside of the cavity, beyond the surface, and note that for high Q modes there is very little energy going outward across the surface compared to the energy inside the cavity, and thus $|U_{\mathrm{E}} - U_{\mathrm{H}}| \ll U$.

With these arguments, we see that for circular symmetry, Eq. F.1 is upheld. Thus, if the electromagnetic energy comes in quanta of $\hbar\omega$, the angular momentum will come in quanta of $m\hbar$.

F.2.3   A Curious Integral

For a circular, conducting cylinder, the integration $U_E = d \int u_E \, dA$ can be done with the help of the relation [8]

$$\int_0^1 y J_m^2(x_{mn} y) dy = \frac{1}{2} J_{m+1}^2(x_{mn}),$$ (F.21)

where x_{mn} is the n-th zero of J_m. Setting the integration $U_H = d \int u_H \, dA$ equal to U_E results in the rather unusual relation

$$\int_0^1 y[J_{m-1}^2(x_{mn} y) + J_{m+1}^2(x_{mn} y)] dy = J_{m+1}^2(x_{mn}).$$ (F.22)

This equation does not seem to be derivable from (F.21) by recursion relations and integration by parts. It also eludes Mathematica [118] and does not appear to be in the common handbooks. Numerical checks of the integral using Mathematica with 16 digit precision show excellent agreement (relative error less than 2×10^{-13}) for the 30 (m, n) combinations with $0 \le m \le 5$, $1 \le n \le 5$.

Appendix G

PARABOLIC CYLINDER FUNCTIONS

The parabolic cylinder functions, $D_p(z)$, are solutions to the differential equation

$$\frac{d^2 u}{dz^2} + \left(p + \frac{1}{2} - \frac{z^2}{4}\right)u = 0. \tag{G.1}$$

All of the relations given here are from Ref. [48].

The definition of D_p in terms of the confluent hypergeometric functions $_1F_1(a; b; z)$ is

$$D_p(z) = 2^{p/2} e^{-z^2/4} \sqrt{\pi} \left(\frac{_1F_1(-\frac{p}{2}; \frac{1}{2}; \frac{z^2}{2})}{\Gamma(\frac{1-p}{2})} - \frac{\sqrt{2} z \, _1F_1(\frac{1-p}{2}; \frac{3}{2}; \frac{z^2}{2})}{\Gamma(\frac{-p}{2})}\right). \tag{G.2}$$

If $|z| \gg 1$, $|z| \gg |p|$, and $|\arg z| < 3\pi/4$, the asymptotic expansion of D_p is

$$D_p(z) \approx e^{-z^2/4} z^p \left(1 - \frac{p(p-1)}{2z^2} + \frac{p(p-1)(p-2)(p-3)}{2 \cdot 4z^4} - \cdots\right) \tag{G.3}$$

Thus, $D_p(z)$ goes to 0 with "Gaussian" speed as $|z| \to \infty$. However, the combination $\exp(z^2/4)D_p(z)$, which is found in our expressions for the Goos-Hänchen shift, can go to ∞, 1, or 0, depending on p.

If p is a nonnegative integer n, one has

$$D_n(z) = (-1)^n e^{z^2/4} \frac{d^n}{dz^n} e^{-z^2/2},$$
$$= 2^{-n/2} e^{-z^2/4} H_n(z/\sqrt{2}). \tag{G.4}$$

The recursion relations are

$$D_{p+1}(z) - zD_p(z) + pD_{p-1}(z) = 0, \tag{G.5}$$

$$\frac{\mathrm{d}}{\mathrm{d}z}D_p(z) - \frac{z}{2}D_p(z) + D_{p+1}(z) = 0. \tag{G.6}$$

To obtain Eq. (E.24), Horowitz and Tamir [116] use the integral relation

$$\int_0^\infty x^{\nu-1}e^{-(\beta/2)x^2-\gamma x}\,\mathrm{d}x = (\beta)^{-\nu/2}\Gamma(\nu)\exp\left(\frac{\gamma^2}{4\beta}\right)D_{-\nu}\left(\frac{\gamma}{\sqrt{\beta}}\right), \tag{G.7}$$

$$(\mathrm{Re}\,\beta > 0, \mathrm{Re}\,\nu > 0)$$

in combination with the linear relation

$$D_p(z) = \frac{\Gamma(p+1)}{\sqrt{2\pi}}[e^{ip\pi/2}D_{-p-1}(iz) + e^{-ip\pi/2}D_{-p-1}(-iz)]. \tag{G.8}$$

Some aspects of computation of parabolic cylinder functions are discussed in Appendix D of Ref. [78].

Appendix H

THE DEBYE EXPANSION AND THE RESONANCE CONDITION

This appendix serves as a reference for Chapter 7.

Straightforward calculations of $U(s)$ involve numerical root finding, and calculations of $U'(s)$ ostensibly involve a numerical derivative which samples via the root finding process. In Sec. 7.4.2 we take derivatives of Bessel functions with respect to their order, a step which again requires a numerical derivative. It would be conceptually helpful to have analytic estimates for these calculations. A small amount of progress can be made in this direction, which we explain here.

As we are typically interested in the case of $\mathrm{Re}(k)R \gg 1$, we would like to use asymptotic expansions for $J_\nu(z)$ and $H_\nu^{(1)}(z)$ which are valid for large z and large ν. The most commonly used large-argument asymptotic expansions for the Bessel functions require $|z| \gg |\nu|$, a condition which does not hold in the present case. There is however, a set of asymptotic expansions that usually are valid. These expansions are known as the Debye expansions, or "approximation by tangents", and are given in [48]. These expansions are strictly valid for positive real order and argument. We will only be using the first terms of each expansion, and numerical checking has revealed that, for the parameter regimes of interest, the approximations seem to be valid when the arguments and/or orders have small imaginary parts. (To determine this validity, one checks the numerical derivatives with respect to order and argument.)

The leading order expansions for $\nu > z$ are

$$J_\nu(z) \approx \frac{\exp[(\nu^2 - z^2)^{1/2} - \nu \operatorname{arccosh}(\nu/z)]}{\sqrt{2\pi}(\nu^2 - z^2)^{1/4}},$$

$$Y_\nu(z) \approx -\sqrt{\frac{2}{\pi}} \frac{\exp[\nu \operatorname{arccosh}(\nu/z) - (\nu^2 - z^2)^{1/2}]}{(\nu^2 - z^2)^{1/4}}, \tag{H.1}$$

which are useful for estimating $H_{UR}^{(1,2)}(kR)$. For $z > \nu$ the leading order expansion of the Bessel functions are

$$J_\nu(z) \approx \sqrt{\frac{2}{\pi}}(z^2 - \nu^2)^{-1/4} \cos\left[(z^2 - \nu^2)^{1/2} - \nu \arccos\left(\frac{\nu}{z}\right) - \frac{\pi}{4}\right],$$

$$Y_\nu(z) \approx \sqrt{\frac{2}{\pi}}(z^2 - \nu^2)^{-1/4} \sin\left[(z^2 - \nu^2)^{1/2} - \nu \arccos\left(\frac{\nu}{z}\right) - \frac{\pi}{4}\right]. \tag{H.2}$$

Ref. [48] warns that the latter two approximations are not valid when $z - \nu$ is of a size comparable to $z^{1/3}$. This condition is much weaker than the $|z| \gg |\nu|$ condition necessary for the more common asymptotic approximation.

Unfortunately, these approximate expressions are transcendental functions, and it is not possible to find a closed form expression for U given k, n, and R. The expansions do however help in the estimation of derivatives with respect to order, which have several uses, including the calculation of $U'(s)$ once a solution (U, k) to the resonance condition has been found.

Using (H.1) gives, for $\nu > z$,

$$\frac{\partial}{\partial \nu} H_\nu^{(1)}(z) \approx -\frac{\nu}{2(\nu^2 - z^2)} H_\nu^{(1)}(z) - \operatorname{arccosh}\left(\frac{\nu}{z}\right) H_\nu^{(2)}(z), \tag{H.3}$$

and using (H.2) gives, for $\nu < z$,

$$\frac{\partial}{\partial \nu} J_\nu(z) \approx \frac{\nu}{2(z^2 - \nu^2)} J_\nu(z) + \arccos\left(\frac{\nu}{z}\right) Y_\nu(z). \tag{H.4}$$

Appendix I

ADDITIONAL PROPERTIES OF THE ELLIPSE

This appendix gives properties of the ellipse that are not mentioned in the main text.

In addition to the "focal point radial" parametrization $r_f(\theta_f)$, an ellipse can be analytically parametrized in regular polar coordinates, $r(\theta)$, where the origin is at the center of the ellipse. This parametrization is given by

$$r(\theta) = a\sqrt{\frac{1 - e^2}{1 - e^2 \cos^2\theta}} = \frac{b}{\sqrt{1 - e^2\cos^2\theta}}. \tag{I.1}$$

Using $ds = \sqrt{r^2 + (dr/d\theta)^2}$, yields

$$\frac{ds}{d\theta} = b\sqrt{\frac{1 - e^2(2 - e^2)\cos^2\theta}{(1 - e^2\cos^2\theta)^3}}. \tag{I.2}$$

When θ_f is taken to be in $(-\pi, \pi]$, θ is given by the geometrical relation

$$\theta(\theta_f) = \arctan\left[\frac{(1 - e^2)\sin\theta_f}{e + \cos\theta_f}\right] + \begin{cases} 0, & |\theta_f| < \pi - \arccos e \\ \pi, & \theta_f > \pi - \arccos e \\ -\pi, & \theta_f < -\pi + \arccos e \end{cases}. \tag{I.3}$$

The inverse, for $\theta \in (-\pi, \pi]$, is given by

$$\theta_f(\theta) = \text{sgn}(\theta) \arccos\left[\frac{-e\tan^2\theta + \text{sgn}\left(\frac{\pi}{2} - |\theta|\right)(1-e^2)^2\sqrt{1 + \frac{\tan^2\theta}{1-e^2}}}{\tan^2\theta + (1-e^2)^2}\right]. \qquad (\text{I.4})$$

References

[1] H. E. Tureci, H. G. L. Schwefel, A. D. Stone, and E. E. Narimanov. Gaussian-optical approach to stable periodic orbit resonances of partially chaotic dielectric micro-cavities. *Opt. Expr.*, 10:752–776, 2002.

[2] J. H. Van Vleck. The correspondence principle in the statistical interpretation of quantum mechanics. *Proc. Natl. Acad. Sci. USA*, 14:178–188, 1928.

[3] A. M. Ozorio de Almeida. *Hamiltonian Systems: Chaos and Quantization*. Cambridge University Press, Cambridge, 1988.

[4] M. Brack and R. K. Bhaduri. *Semiclassical Physics*. Addison-Wesley, Reading, 1997.

[5] P. T. Leung, S. Y. Liu, and K. Young. Completeness and orthogonality of quasinormal modes in leaky optical cavities. *Phys. Rev. A*, 49:3057–3067, 1994.

[6] E. S. C. Ching, P. T. Leung, A. Maassen van den Brink, W. M. Suen, S. S. Tong, and K. Young. Quasinormal-mode expansion for waves in open systems. *Rev. Mod. Phys.*, 70:1545–1554, 1998.

[7] J. U. Nöckel and R. K. Chang. 2-d Microcavities: Theory and Experiments. In R. D. van Zee and J. P. Looney, editors, *Cavity-Enhanced Spectroscopies*, volume 40 of *Experimental Methods in the Physical Sciences*, pages 185–226. Academic Press, San Diego, 2002.

[8] J. D. Jackson. *Classical Electrodynamics, Third Edition*. John Wiley & Sons, Inc., New York, 1999.

[9] B. E. A. Saleh and M. C. Teich. *Fundimentals of Photonics*. John Wiley & Sons, Inc., New York, 1991.

[10] M. Aziz, J. Pfeiffer, and P. Meissner. Modal Behaviour of Passive, Stable Microcavities. *Phys. Stat. Sol. (a)*, 188:979–982, 2001.

[11] F. M. Matinaga, A. Karlsson, S. Machida, Y. Yamamoto, T. Suzuki, Y. Kadota, and M. lkeda. Low-threshold operation of hemispherical microcavity single-quantum-well lasers at 4 K. *Appl. Phys. Lett.*, 62:443–445, 1993.

[12] R. J. Thompson, Q. A. Turchette, O. Carnal, and H. J. Kimble. Nonlinear spectroscopy in the strong-coupling regime of cavity QED. *Phys. Rev. A*, 57:3084–3104, 1998.

[13] P. Maunz, T. Puppe, I. Schuster, N. Syassen, P. W. H. Pinske, and G. Rempe. Normal-Mode Spectroscopy of a Single-Bound-Atom-Cavity System. *Phys. Rev. Lett.*, 94:033002, 2005.

[14] J. A. Sauer, K. M. Fortier, M. S. Chang, C. D. Hamley, and M. S. Chapman. Cavity QED with optically transported atoms. *Phys. Rev. A*, 69:051804, 2004.

[15] M. G. Raymer. colloquium, University of Oregon, 4/24/03, and private communication.

[16] A. Zrener, L. V. Butov, M. Hagn, G. Abstreiter, G. Böhm, and G. Weimann. Quantum Dots Formed by Interface Fluctuations in AlAs/GaAs Coupled Quantum Well Structures. *Phys. Rev. Lett.*, 72:3382–3385, 1994.

[17] D. Gammon, E. S. Snow, B. V. Shanabrook, D. S. Katzer, and D. Park. Fine Structure Splitting in the Optical Spectra of Single GaAs Quantum Dots. *Phys. Rev. Lett.*, 76:3005–3008, 1996.

[18] J. U. Nöckel, G. Bourdon, E. Le Ru, R. Adams, I. Robert, J-M. Moison, and I. Abram. Mode structure and ray dynamics of a parabolic dome microcavity. *Phys. Rev. E*, 62:8677–8699, 2000.

[19] P. W. Milonni and J. H. Eberly. *Lasers*. John Wiley & Sons, New York, 1988.

[20] G. P. Bava, P. Debernardi, and L. Fratta. Three-dimensional model for vectorial fields in vertical-cavity surface-emitting lasers. *Phys. Rev. A*, 63:023816, 2001.

[21] M. Achtenhagen, A. Hardy, and E. Kapon. Three-dimensional analysis of mode discrimination in vertical-cavity surface-emitting lasers. *Appl. Opt.*, 44:2832–2838, 2005.

[22] L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman. Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes. *Phys. Rev. A*, 45:8185–8189, 1992.

[23] M. Padgett, J. Courtial, and L. Allen. Light's orbital angular momentum. *Phys. Today*, 57:35–40, 2004.

[24] D. H. Foster and J. U. Nöckel. Methods for 3-D vector microcavity problems involving a planar dielectric mirror. *Opt. Commun.*, 234:351–383, 2004.

[25] D. H. Foster and J. U. Nöckel. Spatial and polarization structure in micro-domes: effects of a Bragg mirror. In Alexis V. Kudryashov and Alan H. Paxton, editors, *Resonators and Beam Control VII*, volume 5333 of *Proceedings of SPIE*, pages 195–203, 2004.

[26] D. H. Foster and J. U. Nöckel. Bragg-induced orbital angular-momentum mixing in paraxial high-finesse cavities. *Opt. Lett.*, 29:2788–2790, 2004.

[27] A. E. Siegman. *Lasers*. University Science Books, Sausalito, CA, 1986.

[28] P. Yeh. *Optical waves in layered media*. Wiley, New York, 1988.

[29] A. H. Barnett. *Dissipation in Deforming Chaotic Billiards*. PhD thesis, Harvard University, 2000.

[30] D. Cohen, N. Lepore, and E. J. Heller. Consolidating boundary methods for finding the eigenstates of billiards. *J. Phys. A: Math. Gen.*, 37:2139–2161, 2004.

[31] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing, second ed.* Cambridge Univeristy Press, Cambridge, 1992.

[32] F. M. Kahnert. Numerical methods in electromagnetic scattering theory. *J. Quant. Spectrosc. Radiat. Transf.*, 79-80:775–824, 2003.

[33] M. V. Berry. Evanescent and real waves in quantum billiards and gaussian beams. *J. Phys. A*, 27:L391–L398, 1994.

[34] H. E. Türeci. *Wave Chaos in Dielectric Resonators: Asymptotic and Numerical Approaches*. PhD thesis, Yale University, 2003.

[35] H. G. L. Schwefel. *Directionality and Vector Resonances of Regular and Chaotic Dielectric Microcavities*. PhD thesis, Yale University, 2004.

[36] C. W. Erickson. High order modes in a spherical Fabry-Perot resonator. *IEEE Trans. Microwave Theory Tech.*, 23:218–223, 1975.

[37] V. S. Liberman and B. Ya. Zel'dovich. Spin-orbit interaction of a photon in an inhomogeneous medium. *Phys. Rev. A*, 46:5199–5207, 1992.

[38] K. Yu. Bliokh and D. Yu. Frolov. Spin-orbit interaction of photons and fine splitting of levels in ring dielectric resonator. *Opt. Commun.*, 250:321–327, 2005.

[39] M. Lax, W. H. Louisell, and W. B. McKnight. From Maxwell to paraxial wave optics. *Phys. Rev. A*, 11:1365–1370, 1975.

[40] O. Steuernagel. Equivalence between focused paraxial beams and the quantum harmonic oscillator. *Am. J. Phys.*, 73:625–629, 2005.

[41] M. W. Beijersbergen, L. Allen, H. E. L. O. van der Veen, and J. P. Woerdman. Astigmatic laser mode converters and transfer of orbital angular momentum. *Opt. Commun.*, 96:123–132, 1993.

[42] E. Abramochkin and V. Volostnikov. Beam transformations and nontransformed beams. *Opt. Commun.*, 83:123–135, 1991.

[43] H. Laabs and A. T. Friberg. Nonparaxial eigenmodes of stable resonators. *IEEE J. Quantum Electron.*, 35:198–207, 1999.

[44] A. L. Cullen and P. K. Yu. Complex source-point theory of the electromagnetic open resonator. *Proc. R. Soc. London Ser. A*, 366:155–171, 1979.

[45] P. K. Yu and K-M Luk. High order azimuthal modes in the open resonator. *Electron. Lett.*, 19:539–541, 1983.

[46] P. K. Yu and K-M Luk. Field patterns and resonant frequencies of high-order modes in an open resonator. *IEEE Trans. Microwave Theory Tech.*, 32:641–645, 1984.

[47] K-M Luk. Improvement in the resonant formula of a spherical Fabry-Perot resonator with unequal mirrors. *J. Opt. Soc. Am. A*, 3:3–6, 1986.

[48] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products.* Academic Press, Inc., San Diego, 1980.

[49] C. W. Erickson. Perturbation theory generalized to arbitrary (p, l) modes in a Fabry-Perot resonator. *IEEE Trans. Microwave Theory Tech.*, 25:958, 1977.

[50] L. Fratta, P. Debernardi, G. P. Bava, C. Degen, J. Kaiser, I. Fischer, and W. Elsäßer. Spatially inhomogeneously polarized transverse modes in vertical-cavity surface-emitting lasers. *Phys. Rev. A*, 64:031803(R), 2001.

[51] A. M. Sarangan and G. M Peake. Enhancement of Lateral Mode Discrimination in Broad-Area VCSELs Using Curved Bragg Mirrors. *J. Lightwave Tech.*, 22:543–549, 2004.

[52] D. G. Grier. A revolution in optical manipulation. *Nature*, 424:810–815, 2003.

[53] W. M. Lee, X.-C. Yuan, and D. Y. Tang. Optical tweezers with multiple optical forces using double-hologram interference. *Opt. Expr.*, 11:199–207, 2003.

[54] C. Li. Four-photon, five-dimensional entanglement for quantum communication. *Phys. Lett. A*, 313:389–392, 2003.

[55] A. Vaziri, G. Weihs, and A. Zeilinger. Experimental two-photon, three- dimensional entanglement for quant um communication. *Phys. Rev. Lett.*, page 240401, 2002.

[56] A. Mair, A. Vaziri, G. Weihs, and A. Zeilinger. Entanglement of the orbital angular momentum states of photons. *Nature*, 412:313–316, 2001.

[57] E. J. Galvez, P. R. Crawford, H. I. Sztul, M. J. Pysher, P. J. Haglin, and R. E. Williams. Geometric Phase Associated with Mode Transformations of Optical Beams Bearing Orbital Angular Momentum. *Phys. Rev. Lett.*, 90:203901, 2003.

[58] H. Sasada and M. Okamoto. Transverse-mode beam splitter of a light beam and its application to quantum cryptography. *Phys. Rev. A*, 68:012323, 2003.

[59] J. J. Sakurai. *Modern Quantum Mechanics, Second Edition.* Addison-Wesley, Reading, MA, 1994.

[60] H. E. Tureci and A. D. Stone. Deviation from snell's law for beams transmitted near the critical angle: application to microcavity lasers. *Opt. Lett.*, 27:7–9, 2002.

[61] M. C. Gutzwiller. Periodic orbits and classical quantization conditions. *J. Math. Phys.*, 12:343–358, 1971.

[62] M. V. Berry and M. Tabor. Closed orbits and the regular bound spectrum. *Proc. R. Soc. London Ser. A*, 349:101–123, 1976.

[63] H. Poincaré. *Les methodes nouvelles de la mechanique celeste*. Gauthier-Villars, Paris, 1899.

[64] E. J. Heller. Bound-state eigenfunctions of classically chaotic Hamiltonian systems: scars of periodic orbits. *Phys. Rev. Lett.*, 53:1515–1518, 1984.

[65] S. Tomsovic and E. J. Heller. Semiclassical construction of chaotic eigenstates. *Phys. Rev. Lett.*, 70:1405–1408, 1993.

[66] J. U. Nöckel, A. D. Stone, G. Chen, H. Grossman, and R. K. Chang. Directional emission from asymmetric resonant cavities. *Opt. Lett.*, 21:1609–1611, 1996.

[67] S. Lacey, H. Wang, D.H. Foster, and J.U. Nöckel. Directional tunnel escape from nearly spherical optical resonators. *Phys. Rev. Lett.*, 91:033902, 2003.

[68] J. V. José and E. J. Saletan. *Classcal Dynamics: A Contemporary Approach*. Cambridge University Press, Cambridge, 1998.

[69] H. Goldstein, C. Poole, and J. Safko. *Classical Mechanics, Third Edition*. Addison-Wesley, San Francisco, 2002.

[70] L. D. Landau and E. M. Lifshitz. *Mechanics, Third Edition*. Butterworth-Heinenann, Oxford, 1997.

[71] V. I. Arnold. *Mathematical Methods of Classical Mechanics, Second Edition*. Springer-Verlag, New York, 1989.

[72] A. J. Lichtenberg and M. A. Lieberman. *Regular and Chaotic Dynamics, Second Edition*. Springer-Verlag, New York, 1992.

[73] M. Tabor. *Chaos and Integrability in Nonlinear Dynamics*. John Wiley and Sons, New York, 1989.

[74] F. Haake. *Quantum Signatures of Chaos*. Springer-Verlag, Berlin, 1992.

[75] J. U. Nöckel. *Resonances in Nonintegrable Open Systems*. PhD thesis, Yale University, 1997.

[76] S. M. Lacey. *Ray and wave dynamics in three dimensional asymmetric optical resonators*. PhD thesis, University of Oregon, 2003.

[77] M. Hentschel and J. U. Nöckel. The sequential-reflection model in deformed dielectric cavities. In D. Lenstra, T. D. Visser, and K. A. H. van Leeuwen, editors, *Quantum Optics of Small Structures*. Edita KNAW, Amsterdam, 2000.

[78] M. Hentschel. *Mesoscopic wave phenomena in electronic and optical ring structures*. PhD thesis, Max-Planck Institute for the Physics of Complex Systems, 2001.

[79] M. Hentschel and H. Schomerus. Fresnel laws at curved dielectric interfaces of microresonators. *Phys. Rev. E*, 65:045603, 2002.

[80] T. Tamir. Nonspecular phenomena in beam fields reflected by multilayered media. *J. Opt. Soc. Am. A*, 3:558–565, 1986.

[81] W. Nasalski. Longitudinal and transverse effects of nonspecular reflection. *J. Opt. Soc. Am. A*, 13:172–181, 1996.

[82] K. V. Artmann. Berechung der Seitenversetzung des totalreflektierten Strahles. *Ann. Phys. Leipzig*, 2:87–102, 1948.

[83] F. Goos and H. Hänchen. Ein neuer and fundamentaler Versuch zur total Reflection. *Ann. Phys. Leipzig*, 1:333–345, 1947.

[84] R. H. Renard. Total reflection: a new evaluation of the Goos-Hänchen Shift. *J. Opt. Soc. Am.*, 54:1190–1197, 1964.

[85] H. M. Lai, F. C. Cheng, and W. K. Tang. Goos-Hänchen effect around and off the critical angle. *J. Opt. Soc. Am. A*, 3:550–557, 1986.

[86] S. Zhang and C. Fan. Nonspecular phenomena on Gaussian beam reflection at dielectric interfaces. *J. Opt. Soc. Am. A*, 5:1407–1409, 1988.

[87] B. A. Muzykantskii and D. E. Khmelnitskii. Nearly localized states in weakly disordered conductors. *Phys. Rev. B*, 51:5480–5483, 1995.

[88] G. Casati, G. Maspero, and D. L. Shepelyansky. Relaxation process in a regime of quantum chaos. *Phys. Rev. E*, 56:R6233–R6236, 1997.

[89] O. Zaitsev. Quasiclassical Born-Oppenheimer approximations. *Found. Phys.*, 31:7–26, 2001.

[90] O. A. Zaitsev. *Semiclassical surface of section perturbation theory*. PhD thesis, University of Maryland, 2001.

[91] V. M. Babic and V. S. Buldyrev. *Short-wavelength diffraction theory*. Springer Verlag, Berlin, 1972.

[92] J. Mathews and R. L. Walker. *Mathematical Methods of Physics, Second Edition*. Addison-Wesley, Redwood City, CA, 1970.

[93] J. Kevorkian and J. D. Cole. *Multiple Scale and Singular Perturbation Methods*. Springer-Verlag, New York, 1996.

[94] E. Ott. Goodness of ergodic adiabatic invariants. *Phys. Rev. Lett.*, 42:1628–1631, 1979.

[95] M. Wilkinson. Dissipation by identical oscillators. *J. Phys. A: Math. Gen.*, 23:3603–3611, 1990.

[96] C. Jarzynski. Multiple-time-scale approach to ergodic adiabatic systems: Another look. *Phys. Rev. Lett.*, 71:839–842, 1993.

[97] M. V. Berry and J. M. Robbins. Chaotic classical and half-classical adiabatic reactions: geometric magnetism and deterministic friction. *Proc. R. Soc. Lond. A*, 442:659–672, 1993.

[98] D. Cohen. Quantum dissipation due to the interaction with chaotic degrees of freedom and the correspondence principle. *Phys. Rev. Lett.*, 82:4951–4954, 1999.

[99] I. Mrkonjic, D. Radic amd V. Dananic, and V. Lopac. Dynamics of some bouncing ball systems. *Prog. Theor. Phys. Suppl.*, 139:381–391, 2000.

[100] V. Y. Bazhenov, M. S. Soskin, and M. V. Vasnetsov. Screw dislocations in light wavefronts. *J. Mod. Opt.*, 39:985–990, 1999.

[101] S. A. Kennedy, M. J. Szabo, H. Teslow, J. Z. Porterfield, and E. R. I. Abraham. Creation of Laguerre-Gaussian laser modes using diffractive optics. *Phys. Rev. A*, 66:043801, 2002.

[102] D. Akamatsu and M. Kozuma. Coherent transfer of orbital angular momentum from an atomic system to a light field. *Phys. Rev. A*, 67:023803, 2003.

[103] M. P. MacDonald, L. Paterson, K. Volke-Sepulveda, J. Arlt, W. Sibbett, and K. Dholakia. Creation and manipulation of three-dimensional optically trapped structures. *Science*, 296:1101–1103, 2002.

[104] H. Sasada and M. Okamoto. Erratum: Transverse-mode beam splitter of a light beam and its application to quantum cryptography. *Phys. Rev. A*, 69:039901(E), 2004.

[105] R. Simon and G. S. Agarwal. Wigner representation of Laguerre-Gaussian beams. *Opt. Lett.*, 25:1313–1315, 2000.

[106] G. Nienhuis and L. Allen. Paraxial wave optics and harmonic oscillators. *Phys. Rev. A*, 48:656–665, 1993.

[107] S. J. Van Enk and G. Nienhuis. Commutation rules and eigenvalues of spin and orbital angular momentum of radiation fields. *J. Mod. Opt.*, 41:963–977, 1994.

[108] S. M. Barnett and L. Allen. Orbital angular momentum and nonparaxial light beams. *Opt. Commun.*, 110:670–678, 1994.

[109] L. Allen, V. E. Lembessis, and M. Babiker. Spin-orbit coupling in free-space Laguerre-Gaussian light beams. *Phys. Rev. A*, 53:R2937–R2939, 1996.

[110] R. Loudon. Theory of the forces exerted by Laguerre-Gaussian light beams on dielectrics. *Phys. Rev. A*, 68:013806, 2003.

[111] L. Allen, M. J. Padgett, and M. Babiker. The Orbital Angular Momentum of Light. In E. Wolf, editor, *Progress in Optics*, volume 39, pages 291–372. Elsevier, Amsterdam, 1999.

[112] A. T. O'Neil, I. MacVicar, L. Allen, and M. J. Padgett. Intrinsic and extrinsic nature of the orbital angular momentum of a light beam. *Phys. Rev. Lett.*, 88:053601, 2002.

[113] L. Mandel and E. Wolf. *Optical Coherence and Quantum Optics*. Cambridge University Press, Cambridge, 1994.

[114] E. Argence and T. Kahan. *Theory of Waveguides and Cavity Resonators*. Hart Publishing Company, Inc., New York, 1967.

[115] A. Yahalom and Y. Pinhasi amd Y. Lurie. Spectral and variational principles of electromagnetic field excitation in wave guides. *Phys. Lett. A*, 344:18–28, 2005.

[116] B. R. Horowitz and T. Tamir. Lateral displacement of a light beam at a dielectric interface. *J. Opt. Soc. Am.*, 61:586–594, 1971.

[117] R. E. Collin. *Foundations for Microwave Engineering, Second Edition*. McGraw-Hill, New York, 1992.

[118] Wolfram Research, Inc. *Mathematica, Version 5.1*. Champaign, IL, 2003.