

2013

Amplified total internal reflection at the surface of gain medium

Josh Orndorff
The University of Toledo

Follow this and additional works at: <http://utdr.utoledo.edu/theses-dissertations>

Recommended Citation

Orndorff, Josh, "Amplified total internal reflection at the surface of gain medium" (2013). *Theses and Dissertations*. 165.
<http://utdr.utoledo.edu/theses-dissertations/165>

This Thesis is brought to you for free and open access by The University of Toledo Digital Repository. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of The University of Toledo Digital Repository. For more information, please see the repository's [About page](#).

A Thesis

entitled

Amplified Total Internal Reflection at the Surface of Gain Medium

by

Josh Orndorff

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the
Masters of Science Degree in Physics

Dr. Victor Karpov, Committee Chair

Dr. Brian Bagley, Committee Member

Dr. Robert Deck, Committee Member

Dr. Patricia R. Komuniecki, Dean
College of Graduate Studies

The University of Toledo

May 2013

An Abstract of
Amplified Total Internal Reflection at the Surface of Gain Medium

by
Josh Orndorff

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the
Masters of Science Degree in Physics

The University of Toledo
May 2013

Total Internal Reflection (TIR) is the phenomenon whereby a light wave incident on a boundary is completely reflected when the wave's incidence angle exceeds a critical angle, θ_c . For decades there has been debate about whether amplified TIR from a medium exhibiting optical gain is possible, and desire for a theory to explain it. Authors have suggested theories both in favor and in doubt of the phenomenon's existence, and experimental evidence has arose supporting the existence and seemingly simultaneously contradicting proposed theoretical models. In this thesis, reflection coefficients for plane waves are calculated by satisfying boundary conditions from Maxwell's equations at the reflecting surface for both optically lossy and gainy media. Plane wave reflectivity is found to exhibit a discontinuous jump from below unity to above as the incidence angle passes through the critical angle, confirming the existence of amplified TIR. Fourier analysis is used to show that finite beams also exhibit amplified TIR, but do not experience the surprising discontinuous jump in reflectivity at the critical angle.

Acknowledgments

Foremost, I would like to thank Dr. Robert Deck for his unceasing insight, patience, encouragement, and desire to help. I also thank the rest of my committee, Dr. Victor Karpov and Dr. Brian Bagley, for their support of my work. Another thanks goes to my classmates at the university for the time they spent proofing my work, arguing about various physics concepts, and motivating me to keep working hard. In particular, I'd like to thank Joseph Booker for his help with python programming, Aditya Togi, for his help with analytical mathematics, and Alex Cimaroli and Max Junda for their encouragement and helpful suggestions both academically and socially.

I'm also grateful to my family for their support and encouragement throughout my school career. And, finally, I'd like to acknowledge the contributions from the authors, contributors, and supporters of the open-source projects NumPy and Matplotlib. Without their code, I could never have completed the numerical calculations contained herein.

Contents

Abstract	iii
Acknowledgments	iv
Contents	v
List of Figures	vii
List of Abbreviations	ix
List of Symbols	x
1 Background and Motivation	1
1.1 Total Internal Reflection	1
1.2 Lasers and Gain Media	4
1.3 Amplified and Attenuated TIR	5
1.3.1 Amplified TIR with Surface Plasmons	6
1.3.2 Some Interesting Experiments	7
1.3.3 Amplified TIR with Gain Media	8
1.4 Thesis Outline	10
2 Case of Incident Plane Waves	12
2.1 Geometry and Boundary Conditions	12
2.2 Amplitude Reflection Coefficient, r	16
2.3 Intensity Reflection Coefficient, R	17

2.4	Different Polarizations	18
2.5	Different Quadrants for \tilde{k}_{tz}	20
2.5.1	Lossy Medium Below Critical Angle	23
2.5.2	Lossy Medium Beyond Critical Angle	26
2.5.3	Gainy Medium Below Critical Angle	28
2.5.4	Gainy Medium Beyond Critical Angle	32
2.5.5	Conclusions and Experimental Evidence	34
3	Case of a Finite Beam	38
3.1	Derivation and Properties of a Gaussian Beam	39
3.1.1	Coordinate Systems for Reflection	42
3.2	Fourier Analysis of Finite Beams	43
3.2.1	Analytical Method with Approximations	45
3.2.2	Numerical Method – the Discrete Fourier Transform	49
3.2.3	Comparison of the Two Methods	51
3.3	Finite Beam Reflection Profiles Under Many Conditions	54
4	Conclusions	59
	References	61
	A Source Code	63

List of Figures

1-1	Light incident on a lossless boundary above and below the critical angle.	2
1-2	Reflectivity from a lossless boundary.	3
1-3	Internal structure of a laser.	5
1-4	Attenuated reflectivity from a lossy boundary.	6
2-1	Geometry for p polarization.	13
2-2	Geometry for s polarization.	19
2-3	Complex plane diagram for \tilde{k}_{tz}^2	23
2-4	\tilde{k}_{tz}^2 complex plane diagram, $\gamma < 0$, $\theta_i < \theta_c$	24
2-5	Two possible Reflectivities from lossy medium below critical angle	25
2-6	\tilde{k}_{tz}^2 complex plane diagram, $\gamma < 0$, $\theta_i > \theta_c$	26
2-7	Two possible Reflectivities from lossy medium above critical angle	27
2-8	Reflectivity, R , from lossy medium at various losses.	28
2-9	\tilde{k}_{tz}^2 complex plane diagram, $\gamma > 0$, $\theta_i < \theta_c$	29
2-10	Two possible Reflectivities from gainy medium below critical angle	30
2-11	\tilde{k}_{tz}^2 complex plane diagram, $\gamma > 0$, $\theta_i > \theta_c$	32
2-12	Final Reflectivity for Plane Waves – Low Gain	34
2-13	Final Reflectivity for Plane Waves – High Gain	35
2-14	Detailed quadrant summary of \tilde{k}_{tz}^2 and \tilde{k}_{tz} under various conditions	36
3-1	Gaussian beam diagram showing beam waist and Rayleigh range	39
3-2	Cross-section intensity profiles of a Gaussian beam.	40

3-3	Gaussian beam diagram showing phase front radius	42
3-4	Natural axes of incident and reflected Gaussian beams	43
3-5	Comparison of exact and approximated Gaussian beams at the reflecting surface.	48
3-6	DFT schematic illustration.	50
3-7	Comparison of analytical and FFT Fourier transform methods	52
3-8	Intensity profile of incident and reflected Gaussian beams	53
3-9	Finite Beam Reflectivities at various spot sizes (<i>p</i> -polarization)	55
3-10	Finite Beam Reflectivities at various spot sizes (<i>s</i> -polarization)	56
3-11	Finite Beam Reflectivities at Various Gain Factors	57
3-12	Finite Beam Reflectivities at a few incident angles, plotted vs γ	58

List of Abbreviations

TIR	Total Internal Reflection – When a plane wave is completely reflected from a boundary.
aTIR	Attenuated Total Internal Reflection – TIR wherein the reflected beam is less intense than the incident beam.
ATIR	Amplified Total Internal Reflection – TIR wherein the reflected beam is more intense than the incident beam.
DFT	Discrete Fourier Transform.
FFT	Fast Fourier Transform – A time efficient algorithm for calculating the DFT.
GH Shift	Goos-Hanchen shift. The amount by which a reflected beam is shifted across a boundary from that predicted by geometrical optics.
FDTD	Finite difference time domain. A numerical method for solving differential equations including Maxwell’s equations.

List of Symbols

c	Light's speed in free space. $c = 2.998$ m/s
e	Euler's constant. $e \approx 2.71828$
i	The imaginary constant. $i \equiv \sqrt{-1}$
k	Wave number of an electromagnetic wave. $k = n\frac{2\pi}{\lambda} = n\frac{\omega}{c}$
\mathbf{k}	Wave vector. This vector's magnitude is the wave number, and its direction is the wave's propagation direction.
k_i	Wave number specifically for an incident wave. Subscripts r (reflected) and t (transmitted) are also used.
\tilde{k}_{tz}	Complex-valued wave number for a transmitted beam's z -component.
k_m	k -domain's m^{th} sample in a discrete Fourier transform.
n_1	Real-valued refractive index characterizing the medium from which light is incident on a boundary.
\tilde{n}_2	Complex-valued refractive index characterizing the medium into which light is transmitted, or from which it is reflected. $\tilde{n}_2 = n_2 - i\gamma$
n_2	Real part of \tilde{n}_2 . Characterizes the propagation of light in medium two.
\tilde{r}	Amplitude reflection coefficient for wave incident on a boundary. $\tilde{r} \equiv E_r/E_0$
w	Width of a Gaussian beam. The radial distance away from the center of the beam at which the light's intensity has fallen off by a factor of two.
w_0	Waist of a Gaussian beam. The width of the beam at its narrowest point.
\mathbf{x}	Position vector of components x , y , and z , in the 3D space spanned by \hat{x} , \hat{y} , and \hat{z} . The boundary is defined by the plane $z = 0$.
\mathbf{x}'	Position vector in the coordinate system of a Gaussian beam. The beam propagates along the z' direction.
x_n	x -domain's n^{th} sample in a discrete Fourier transform.
\mathbf{D}	Electric displacement vector.
\mathbf{D}_i	Electric displacement field vector of incident plane wave. Subscripts r (reflected) and t (transmitted) are also used.
\mathbf{E}	Electric field vector.

\mathbf{E}_i	Electric field vector of incident plane wave. Subscripts r (reflected) and t (transmitted) are also used.
E_n	n^{th} x -domain electric field sample in a discrete Fourier transform.
F_m	m^{th} k -domain electric field sample in a discrete Fourier transform.
L	Spacial sampling interval in a discrete Fourier transform. $L = 1/\kappa_s$.
N	Total number of samples (both spacial and wave number) in a discrete Fourier transform.
R	Intensity reflection coefficient. $R = \tilde{r}^* \tilde{r}$.
γ	Imaginary part of \tilde{n}_2 . Characterizes the absorption ($\gamma < 0$) or amplification ($\gamma > 0$) of light in medium two.
ϵ_1	Real-valued permittivity characterizing the medium from which light is incident on a boundary.
$\tilde{\epsilon}_2$	Complex-valued permittivity characterizing the medium into which light is transmitted or from which it is reflected. $\tilde{\epsilon}_2 \equiv \epsilon'_2 + i\epsilon''_2$
ϵ'_2	Real part of $\tilde{\epsilon}_2$. Characterizes the propagation of light in medium two.
ϵ''_2	Imaginary part of $\tilde{\epsilon}_2$. Characterizes the absorption ($\epsilon''_2 > 0$) or amplification ($\epsilon''_2 < 0$) of light in medium two.
θ_c	Critical angle of TIR
θ_i	Incidence angle of a plane wave on a boundary.
θ_r	Reflection angle of a plane wave from a boundary.
$\tilde{\theta}_t$	Transmission angle of a plane wave (or Gaussian beam's central ray) from a boundary. $\tilde{\theta}_t$ is complex valued except when no gain or loss is present in medium two.
θ_p	The incident angle necessary to excite a surface plasmon.
κ_s	Spacial sampling wave number in a discrete Fourier transform. $\kappa_s = 1/L$
λ	Wavelength of a plane wave
μ	Magnetic permeability of a medium. In this thesis all media are assumed to be non-magnetic. $\mu = 1$ in Gaussian units or $\mu = \mu_0$ in SI units.
ϕ	Incidence angle of the central ray in a Gaussian beam.
ξ	Gouy phase of a Gaussian beam. $\xi \equiv \arctan z'/z_R$.
ω	Angular frequency of a light wave.

Chapter 1

Background and Motivation

Like all original research, this thesis builds on several ideas that have been previously established in optics. In this introduction, I will summarize the relevant work that has already been done, illustrate and exemplify the debate that has surrounded the single surface amplified total internal reflection phenomenon, and very briefly outline the research presented in subsequent chapters.

1.1 Total Internal Reflection

The phenomenon of total internal reflection (TIR) has been well known for centuries. It can be observed in everyday life and is well understood theoretically[5]. When a wave is incident on a boundary surface, some of its energy, in general, is reflected back from the boundary, and some of its energy is transmitted into the second medium behind the boundary. The fact that boundary conditions exist means there are spacial and temporal constraints that lead to kinematic properties such as the law of reflection which states the reflection angle must be equal to the incident angle, $\theta_i = \theta_r$, and Snell's law which gives the transmission angle as $\sin \theta_t = \frac{n_1}{n_2} \sin \theta_i$, where n_1 and n_2 are the refractive indices of both media. Both of these laws will be re-derived in chapter 2 and will prove to be unaffected by the presence of gain or loss in the second medium.

The angle at which the right hand side of Snell's law exceeds one is known as the critical angle, denoted θ_c . At this angle when TIR first occurs, no wave exists in the second medium whatsoever. Beyond the critical angle, TIR continues to occur, but an evanescent, or exponentially decaying, wave is excited in the second medium. It is this evanescent wave that is responsible for the Goos-Hanchen shift (GH shift)[4].

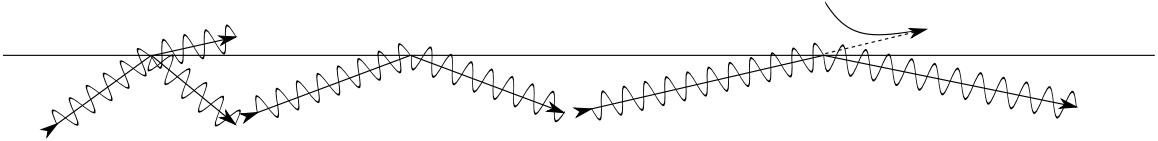


Figure 1-1: Light incident on a lossless boundary above and below the critical angle. Left - Below the critical angle both a reflected and transmitted beam exist. Middle - At the critical angle only a totally reflected beam exists. Right - Beyond the critical angle a totally reflected beam and an evanescent field exist.

In addition to the kinematic properties, the specific nature of the boundary conditions provides a set of dynamic equations, which define certain Fresnel coefficients describing the relative brightness of the reflected and transmitted waves. The Fresnel coefficients for lossless media are derived in most entry level optics texts [5] [6], and will be re-derived in chapter 2. Unlike the law of reflection and Snell's law, we will see that the Fresnel coefficients do change with the presence of gain or loss in medium two (sometimes in unintuitive and surprising ways ex. fig 1-4).

Practical applications of TIR are numerous. In the science lab, Abbe refractometry is used to determine refractive indices. In daily life, communication and data transfer happens largely over fiber optic cables which make use of TIR. Consequently a thorough theoretical understanding of all its manifestations is warranted.

The classical treatment discussed above is sufficient to describe the TIR phenomenon in cases of nearly transparent media, and incident beams that are not narrowly focused. The most useful results such as the critical angle, law of reflection and

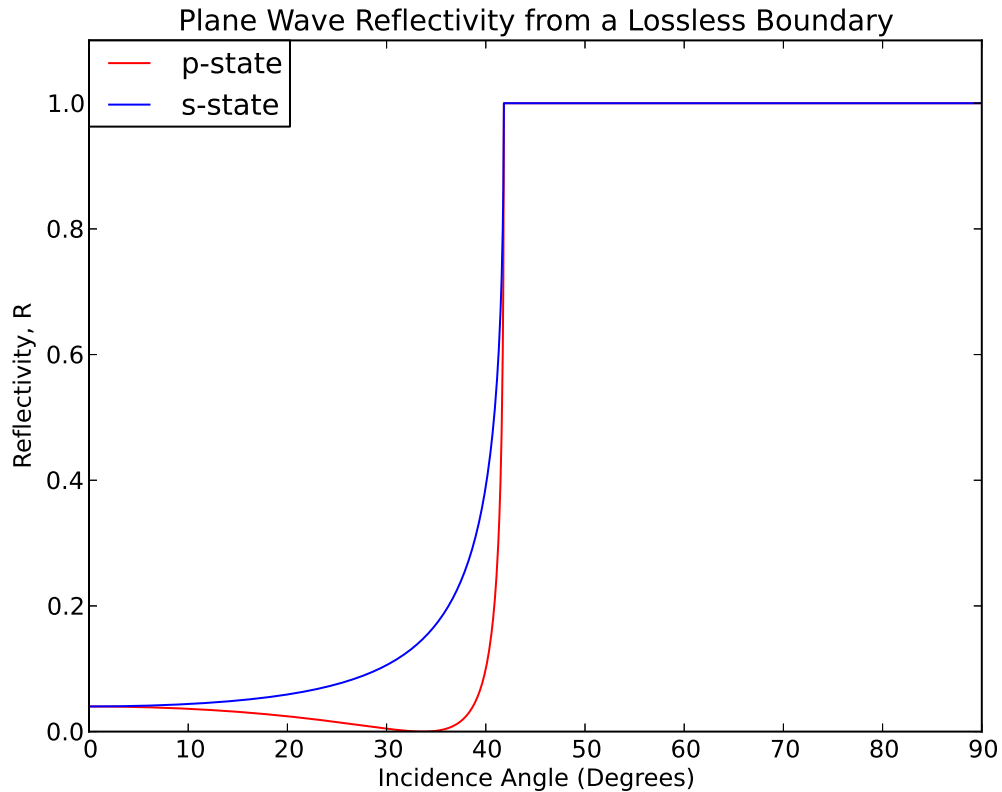


Figure 1-2: The plot shows intensity reflection coefficient, R , for plane waves from a boundary where both media are lossless. Note that the reflectivity is 100% for all angles beyond the critical angle, but never exceeds 100%.

Snell's law are important enough that they are taught as part of most introductory physics courses. But the treatment above does not treat all physical properties of a real-world system, and is therefore incomplete. Our understanding of total internal reflection can be made more robust by extending our model to correct for at least two deficiencies in the usual derivation.

First, the classical treatment assumes all media are perfectly transparent and lossless. In reality the vast majority of media, and certainly the media that we interact with everyday, such as glass, air, water, etc. are all lossy, meaning they absorb some of the light that passes through them. And some media such as laser dyes and other

active media with inverted atomic populations are active or gainy, meaning they amplify the light that passes through them. The lossy or gainy nature of such media can be accounted for by introducing a complex refractive index, $\tilde{n} = n - i\gamma$. This complex refractive index will be used in chapter 2 and throughout the remainder of the thesis. The extent to which these media absorb or amplify light depends on the frequency of light in question, so that in general $\tilde{n} = \tilde{n}(\omega)$, but the theory can be developed for a single frequency with the refractive index, \tilde{n} , taken to be a constant.

Second, the classical treatment assumes that both the incident plane wave and the boundary extend infinitely in directions perpendicular to the incidence plane. More generally, edge effects occur whenever either the beam or the boundary terminates, as they must in the real world. But in practice a laser beam with a width more than a dozen wavelengths incident on a boundary of more than a dozen beam widths can be well approximated as an infinite plane wave. Finite size effects become especially important when the beam width is on the order the incident light's wavelength.[9]

This thesis will begin by re-deriving the Fresnel reflection coefficients for a second medium with gain or loss (focusing primarily on gain), and will discuss the significance of the results, considering the effects of both gain in layer two and finite beam width.

1.2 Lasers and Gain Media

The discovery of stimulated emission led to many advances in our understanding of optics as well as our ability to manufacture useful optical devices. By exciting media in clever ways so that their populations are in a non-thermal state where a majority of atoms are excited, known as population inversion, a gain medium is achieved such that stimulated emission will amplify light that passes through the gain medium.

The laser cavity depicted above results in the emission of a narrowly focused, approximately Gaussian Intensity profile. Without the narrow beam of a laser, many

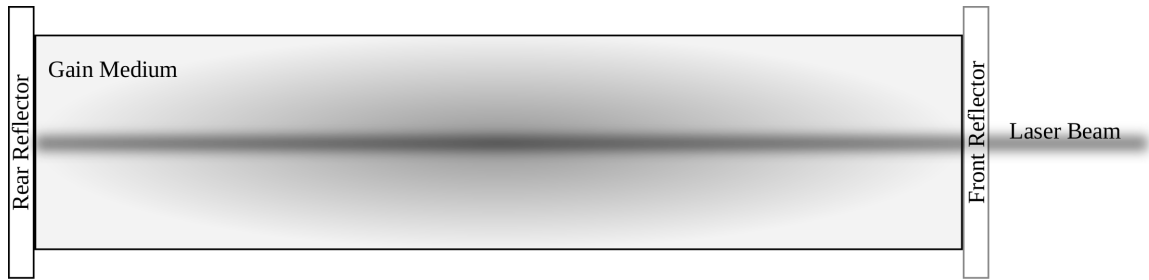


Figure 1-3: As light propagates through the gain medium in a laser, it is amplified via stimulated emission and a laser beam is produced.

of the TIR applications listed previously, as well as the experimental work outlined in section 1.3.2 would be impossible.

This thesis investigates theoretically the case where no wave propagates in a gain medium, but rather an evanescent wave exists in the gainy region. It will also discuss the details of a Gaussian laser beam in chapter 3 and use the formalization to predict real-world experimental results where only a finite beam (such as a Gaussian) can be realized.

1.3 Amplified and Attenuated TIR

Amplified total internal reflection (ATIR) is the phenomenon whereby a wave reflected from a boundary is more intense than the incident wave. In the classical treatment above, this is clearly not possible as it would violate energy conservation. But as we know the treatment above is not the whole story.

Returning to the case of two semi-infinite media, and a plane boundary between them, we consider the case of a lossy second medium. It is established [4] [14] that attenuated total internal reflection occurs in this case, and the reflection coefficients will be derived in chapter 2.

The existence of such attenuated reflection begs the question whether amplified

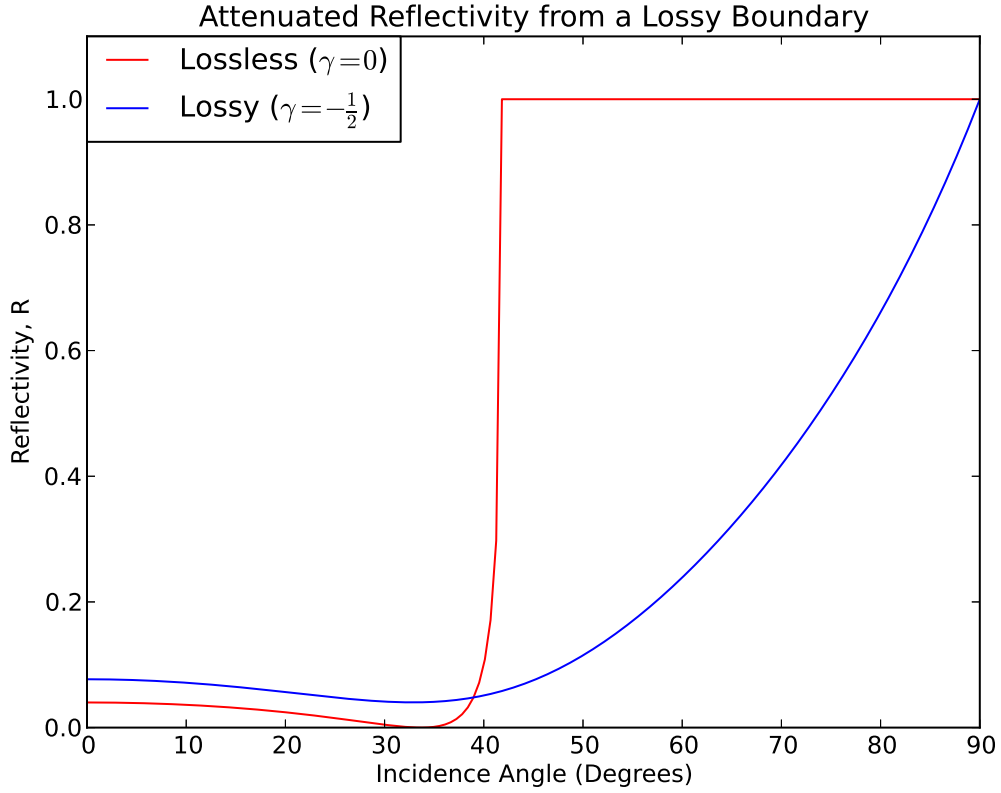


Figure 1-4: P-state intensity reflection coefficients, R , for plane waves from a boundary where both media are lossless (red) and where medium two is lossy (blue). It is worth noting that reflectivity below the critical angle is actually higher in the lossy case despite loss in medium 2.

reflection could exist if the second medium were gainy rather than lossy. The existence of an amplified total internal reflection would serve as a nice counterpart to this attenuated total internal reflection, but as we will see in section 2.5, there is a mathematical curiosity that has caused some authors to doubt its existence.

1.3.1 Amplified TIR with Surface Plasmons

To lend further credence to the idea of ATIR from a single gain medium boundary, it was shown[12] that enhanced reflection is theoretically possible if a thin metal film

is inserted between the two media and the incidence angle is near that necessary to excite a surface plasmon, $\theta_i = \theta_p > \theta_c$.

Using modified Fresnel reflection coefficients similar to those derived in sections 2.2 and 2.3 the authors of [12] show that after the gain exceeds a certain threshold value, reflectivity is resonantly enhanced for incident angles in the vicinity of the plasmon angle, $\theta_i \approx \theta_p$. The authors also estimate that such enhanced reflection could be experimentally achievable in the near infrared using metals and gain media that could be practically manufactured.

1.3.2 Some Interesting Experiments

In 1966 Charles J. Koester of the American Optical Company published a concise report detailing experimental results that he had achieved. His research team had previously had success transmitting and amplifying light through a fiber optic cable with a gainy core [7]. They referred to such devices as “active core fiber lasers”. The devices transmitted light via TIR in the same way as a traditional fiber optic cable, but also amplified the light as it traveled. In a later experiment Koester returned to a traditional passive core, and instead used an active cladding. In this case he again found that light was amplified as it traversed the fiber despite experiencing TIR from the cladding. He interpreted this result to mean that light was reflected from the cladding with reflectivity greater than one. We will see in the following section that some authors have argued that Koester misinterpreted his results, but I will ultimately support his conclusion.

In 1972 a group of Soviet scientists, Kogan, Volkov, and Lebedev, performed an experiment at the Institute of Organic Semiconductors and Dyes showing that ATIR was possible with a single reflection from a planar surface [8]. The report is sparsely quantitative, but they do report an intensity reflectivity of 25. The following year, the same group published a follow-up experiment reporting a single reflection gain of

roughly 1000[10].

These two experiments, both of which demonstrate amplified reflection without surface plasmon coupling appear to further substantiate the intuitive thought that, if lossy media cause attenuated reflection, then gainy media may lead to amplified reflection.

1.3.3 Amplified TIR with Gain Media

Despite the experimental evidence discussed in the previous section, the mathematical results discussed in section 2.5 still lead authors to support drastically different physical interpretations of the ATIR phenomenon.

The first theory attempting to explain amplified total internal reflection came just months after Kogan et. al.'s experiment was published. In September 1972 two Soviet theorists, Romanov and Shakhidzhanov published their work explaining ATIR using complex refractive indices and well known boundary conditions at the reflecting surface [13]. Their work was consistent with the sparse experimental data available at the time, and is the basis for much subsequent theoretical work including this thesis.

Just a few more months after Romanov and Shakhidzhanov published their theory, Kogan et. al. published the results of their second experiment [10]. Their purported single pass gain of roughly 1000 was inconsistent with the recently published theory of Romanov and Shakhidzhanov. Two years later a new theory came from two American authors at the University of Maine. Callary and Carniglia derived a new theory that was distinct from the Romanov and Shakhidzhanov theory in two important ways. First they treated a three layer problem where the gain medium was of finite thickness between two transparent half spaces. Their results regarding the two layer problem came from taking the limit as $d \rightarrow \infty$ for their gain slab. Second, they chose the complex conjugate wave vector for the transmitted beam (discussed in detail in section 2.5) which resulted in gain at all angles both above and below the critical

angle. Their new theory appeared to be in agreement with the new experimental data, and allowed for arbitrarily high reflectivity (even allowing it to be infinite) for the correct combinations of layer thickness and incident angle.

In 1978, Plotz et. al. from the University of Toledo showed theoretically that enhanced reflectivity can be achieved when a thin metal layer is placed between the incident medium and the gain medium if the incidence angle is chosen to excite a surface plasmon on the metal surface [12]. Although treating a somewhat different problem, the results, like those of Callary and Carniglia, allowed for arbitrarily high (even infinite) reflectivity for the appropriate incidence angle and metal layer thickness. These results lend support to the Callary and Carniglia theory because they validate the surprising result of infinite gain.

Decades later in 2003, Fan et. al. revisited the topic of single surface ATIR and built on the work of R and S to calculate the Goos Hanchen shift (GH shift) for plane waves incident on a gainy medium [4]. They made the same argument as presented in the original R and S paper to conclude that ATIR was possible beyond the critical angle but not below it. Although they cited Callary and Carniglia to support their claim that ATIR was possible beyond the critical angle, they concluded that Callary and Carniglia were incorrect in stating that ATIR was possible at all angles. They did not offer an alternative explanation for Kogan et. al.'s measured gain of 1000.

Willis et. al. continued building on the theory of R and S and treated the important case of finite width beams theoretically for the first time [14] in 2008. They elegantly showed both analytically and via finite difference time domain (FDTD) simulations that a finite width beam will exhibit ATIR beyond the critical angle, but will not experience the surprising discontinuity that plane waves do. However they, like Fan et. al., still did not account for the gain of 1000 that Kogan et. al. reported.

In 2011 a manuscript by Tobias and Masud Mansuripur attempted to explain all the experimental data, possibly including Kogan et. al.'s gain of 1000 result, using

an entirely different interpretation of the theory [11]. Taking the exact opposite route from Callary and Carniglia, the Mansuripurs suggested that single surface ATIR was not possible at any angles, and claimed that reflection from lossy or gainy media was always less than one. They argued that the observed amplification was in fact not due to reflection from the front surface of the material, but was observed because in all real experimental setups the gainy region is of finite thickness. They proposed that the observed amplification was due to a transmitted wave propagating through the gainy slab, partially reflecting from its back face, and propagating back to the front face. Their manuscript was not accepted for publication, but was received as part of a private communication on the matter.

The history of this problem is long and interesting, with a revival in the literature in recent years. Proposed theoretical models range from Callary and Carniglia who conclude that ATIR is possible at all incident angles, to the Mansuripurs who claim that it is not possible at any angles. In section 2.5, I will discuss in detail the mathematical results that lead to this debate.

1.4 Thesis Outline

In the next chapter I will set up a standard coordinate system, define the variables and conventions used throughout this thesis, and derive the Fresnel reflection coefficients from Maxwell's equations. I will then discuss and interpret the two possible mathematical solutions that led to the debate previously discussed, and show that the mathematical solution that leads to amplified total internal reflection is the physically correct solution.

In chapter 3, I will build on the results of chapter 2 to treat the reflectivity of finite width laser beams and investigate interesting edge effects that emerge when the beam is only a few wavelengths wide. I will also compare and evaluate the various

numerical and analytical methods for performing such analyses. In the final section, I will consider the effect on the enhanced reflectivity produced by changes in various parameters of the system, such as refractive indices, gain, beam focus, polarization, and incident angle, and will discuss the significance of these results.

Finally, in chapter 4, I will summarize the detailed conclusions from throughout the text and emphasize their importance.

Chapter 2

Case of Incident Plane Waves

In this chapter I establish the geometry and naming conventions followed throughout the thesis and use them to re-derive the Fresnel plane wave reflection coefficients for the case where the second medium is gainy or lossy. I also discuss the slight differences that arise when treating p and s polarization of incident light. Finally in section 2.5 I resolve the various interpretations of the mathematics that have led to dispute about the existence of single surface amplified total internal reflection.

2.1 Geometry and Boundary Conditions

The problem addressed in this thesis is fundamentally a two layer problem; its geometry is depicted in figure 2.1. The boundary lies in the $z = 0$ plane and the incidence plane is given by $y = 0$. The medium on the left is transparent (neither lossy nor gainy) and characterized by the real-valued refractive index n_1 . It is home to the incident wave which strikes the boundary at incident angle θ_i relative to its normal, and the reflected wave which is reflected back into the same medium at a reflection angle θ_r . The medium on the right is home to the transmitted wave which is refracted from the boundary at transmission angle, θ_t and is characterized by complex refractive index, $\tilde{n}_2 = n_2 - i\gamma$. The gain factor is called γ , and the explicit negative sign in the definition means that positive gamma is gainy and negative gamma is

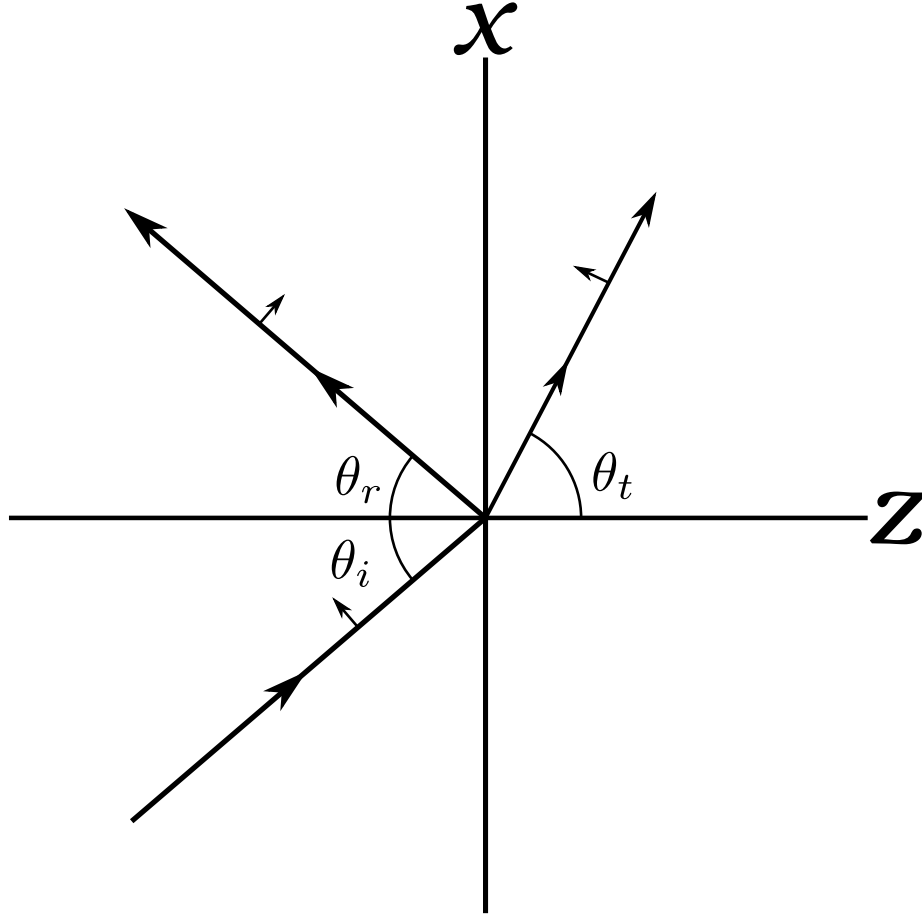


Figure 2-1: Geometry for p polarization. Large arrows represent propagation direction of plane waves. Small arrows are direction of electric fields.

lossy. All quantities that are known to be complex will be explicitly indicated as such by the presence of a tilde over their variables, as is the case with \tilde{n}_2 . Having specified the media's refractive indices, their permittivities can be calculated as $\epsilon_1 = n_1^2$ and $\tilde{\epsilon}_2 = \tilde{n}_2^2$.

In figure 2.1 the long rays indicate the direction in which waves are traveling. As is standard, wave vectors, \mathbf{k} , will point in the direction of propagation of their respective waves, and their magnitudes will be given by $k = n \frac{\omega}{c}$. The short arrows adjacent to each ray indicate the direction of the \mathbf{E} -field for each wave. With these geometry conventions established, the wave vectors and fields of each wave can be

written explicitly.

$$k_i = k_r = n_1 \frac{\omega}{c} \quad (2.1a)$$

$$\tilde{k}_t = \tilde{n}_2 \frac{\omega}{c} = k_t - i\gamma \frac{\omega}{c} \quad (2.1b)$$

Note the definition of k_t (without a tilde) in the immediately preceding equation.

$$\mathbf{E}_i(\mathbf{x}, t) = E_{0i}(\cos \theta_i \hat{\mathbf{x}} - \sin \theta_i \hat{\mathbf{z}}) e^{i(k_{ix}x + k_{iz}z - \omega t)} \quad (2.2)$$

$$\mathbf{E}_r(\mathbf{x}, t) = E_{0r}(\cos \theta_r \hat{\mathbf{x}} + \sin \theta_r \hat{\mathbf{z}}) e^{i(k_{rx}x + k_{rz}z - \omega t)} \quad (2.3)$$

$$\mathbf{E}_t(\mathbf{x}, t) = E_{0t}(\cos \theta_t \hat{\mathbf{x}} - \sin \theta_t \hat{\mathbf{z}}) e^{i(k_{tx}x + k_{tz}z - \omega t)} e^{\gamma \frac{\omega}{c} (\sin \theta_t x + \cos \theta_t z)} \quad (2.4)$$

Here the electric field expressions do not have tildes despite explicit i 's in the exponents. This indicates that the actual electric field is given by the real part of the expression alone. The extra real exponential in \mathbf{E}_t is responsible for the exponential growth or decay that a wave experiences when it propagates through a gainy or lossy medium.

Having established explicit forms for the electric field for each wave, the boundary conditions can be imposed at the surface (the $z = 0$ plane). Before the specific continuity conditions can be satisfied at the boundary, the spacial and temporal constraints must be satisfied. That is, if the specific boundary conditions are to be met at all points and at all times on the boundary, the space and time dependence of each wave must be the same at each point on the boundary. It is clear by investigating equations 2.2 and following that the time dependence is already satisfied and can be subtracted from each term which, at $z = 0$, leaves,

$$k_{ix}x = k_{rx}x = k_{tx}x + \frac{\gamma\omega}{ic} \sin(\theta_t)x \quad (2.5)$$

Equality of the first two terms in equation 2.5 leads to the law of reflection, $\theta_i = \theta_r$.

It comes as no surprise that the incident angle and reflected angle are congruent, as this result is well known in the case of transparent media and is in all introductory physics texts. However the result is still noteworthy as it applies equally well to the more general case of gain or loss in medium two.

Investigating the second two terms in Eq: 2.5 allows us to derive Snell's law, but the gain factor in medium two makes the result less trivial than the case of transparent media.

$$k_i \sin \theta_i = \tilde{k}_t \sin \theta_t \quad (2.6)$$

The left hand side of equation 2.6 is real as always, but the gain factor in medium two has caused the right hand side to be properly complex. In past works this issue has been addressed in two different ways.

First, one can assume that no gain is present exactly at the boundary which restores Snell's law to its usual form. Because the boundary is infinitesimally thin, it is really part of neither medium one nor medium two, so the choice of gain or no gain is arbitrary. This method leads to the correct results but seems to lack mathematical rigor.

Second, one can allow the transmission angle θ_t to be complex-valued. This is my preferred method as it preserves mathematical rigor and doesn't require a no-gain condition at the boundary. It is notable that both methods ultimately yield the same results. Throughout the rest of the thesis, the transmission angle will be written as $\tilde{\theta}_t$, and equation 2.6 will be rewritten as follows.

$$k_i \sin \theta_i = \tilde{k}_t \sin \tilde{\theta}_t \quad (2.6')$$

From this refined version of Snell's law and the identities $k_{ix} = k_i \sin \theta_i$ and $k_{tx} = \tilde{k}_t \sin \tilde{\theta}_t$, we can see that $k_{ix} = k_{tx}$ and therefore both must be real. So, although \tilde{k}_t is complex, the presence of boundary conditions has forced the x -component, k_{tx} to

be real and equal to k_{is} . This result means that the entire imaginary part of \tilde{k}_t must come from its z component, and consequently equation 2.4 can be simplified.

$$\mathbf{E}_t(\mathbf{x}, t) = E_{0t}(\cos \theta_t \hat{\mathbf{x}} - \sin \theta_t \hat{\mathbf{z}}) e^{i(k_{tx}x + k_{tz}z - \omega t)} e^{\gamma \frac{\omega}{c} \cos \theta_t z} \quad (2.4')$$

If one makes the argument that there is no gain exactly at the boundary rather than allowing θ_t to be complex, care must be taken to argue that k_{tx} is real-valued to obtain this result.

2.2 Amplitude Reflection Coefficient, r

Having satisfied the spacial and temporal constraints at the boundary, the appropriate continuity conditions can be applied to the fields at the boundary. This will lead naturally to the amplitude reflection coefficients. It is well established that the parallel component of electric field must be continuous across the boundary, and the normal component of displacement field must be likewise[6].

$$E_{ix} + E_{rx} = E_{tx} \quad (2.7a)$$

$$D_{iz} + D_{rz} = D_{tz} \quad (2.7b)$$

The waves' amplitudes have already been expressed in equations 2.2, 2.3, and 2.4', and conveniently separated into component form. They can be simplified and rewritten using the law of reflection and Snell's law to eliminate all occurrences of θ_r , and some occurrences of $\tilde{\theta}_t$.

$$\mathbf{E}_{0i} = E_{0i}(\cos \theta_i \hat{\mathbf{x}} - \sin \theta_i \hat{\mathbf{z}}) \quad (2.8a)$$

$$\mathbf{E}_{0r} = E_{0r}(\cos \theta_i \hat{\mathbf{x}} + \sin \theta_i \hat{\mathbf{z}}) \quad (2.8b)$$

$$\mathbf{E}_{0t} = E_{0t}(\cos \tilde{\theta}_t \hat{\mathbf{x}} - \frac{n_1}{\tilde{n}_2} \sin \theta_i \hat{\mathbf{z}}) \quad (2.8c)$$

The amplitude reflection coefficient, \tilde{r} , is calculated by substituting the appropriate amplitude components into the two boundary conditions.

$$E_{0i} \cos \theta_i + E_{0r} \cos \theta_i = E_{0t} \cos \tilde{\theta}_t \quad (2.9a)$$

$$E_{0i} + E_{0r} = \frac{\tilde{n}_2}{n_1} E_{0t} \quad (2.9b)$$

Finally, combining equations 2.9 and solving for the ratio $\tilde{r} \equiv E_{0r}/E_{0i}$ yields a functional form for the amplitude reflection coefficient \tilde{r} in terms of known parameters. The expression can be written in many forms, and only the two most useful are listed here. In terms of refractive indices and known angles,

$$\tilde{r} = \frac{n_1 \cos \tilde{\theta}_t - \tilde{n}_2 \cos \theta_i}{n_1 \cos \tilde{\theta}_t + \tilde{n}_2 \cos \theta_i} \quad (2.10)$$

or in terms of permittivities and wave vector components,

$$\tilde{r} = \frac{\epsilon_1 \tilde{k}_{tz} - \tilde{\epsilon}_2 k_{iz}}{\epsilon_1 \tilde{k}_{tz} + \tilde{\epsilon}_2 k_{iz}} \quad (2.11)$$

These equations are the final result of this section. The amplitude reflection coefficients are complex numbers, and can be separated into the form $\tilde{r} = r_r + ir_i$. The amplitude reflection coefficients will be most useful when analyzing the interaction of a Gaussian beam with a gainy boundary in chapter 3.

2.3 Intensity Reflection Coefficient, R

Our task now is to calculate the intensity reflection coefficient, R , for incident plane waves. The algebra is simple as the intensity reflection coefficient is calculated

as $R = \tilde{r}^* \tilde{r}$. At this point it is convenient to give names to the real and imaginary parts of wave vectors and permittivities.

$$\tilde{k}_{tz} = k'_{tz} + ik''_{tz} \quad (2.12)$$

$$\tilde{\epsilon}_{tz} = \epsilon'_{tz} + i\epsilon''_{tz} \quad (2.13)$$

Using these two definitions, the amplitude reflection coefficient can be separated explicitly into real and imaginary parts.

$$\tilde{r} = \frac{(\epsilon_1 k'_{tz} - \epsilon_2' k_{iz}) + i(\epsilon_1 k''_{tz} - \epsilon_2'' k_{iz})}{(\epsilon_1 k'_{tz} + \epsilon_2' k_{iz}) + i(\epsilon_1 k''_{tz} + \epsilon_2'' k_{iz})} \quad (2.14)$$

And the intensity reflection coefficient follows immediately.

$$R = \tilde{r}^* \tilde{r} = \frac{(\epsilon_1 k'_{tz} - \epsilon_2' k_{iz})^2 + (\epsilon_1 k''_{tz} - \epsilon_2'' k_{iz})^2}{(\epsilon_1 k'_{tz} + \epsilon_2' k_{iz})^2 + (\epsilon_1 k''_{tz} + \epsilon_2'' k_{iz})^2} \quad (2.15)$$

The above is in fact the correct expression for R , but it is not entirely in terms of known parameters of the system. It is easily seen from the definition of permittivity that $\epsilon_2' = n_2^2 - \gamma^2$ and $\epsilon_2'' = -2n_2\gamma$. Finding \tilde{k}_{tz} is not as easy, and is discussed in detail in section 2.5. But at this point I will concisely summarize the differences between p- and s- polarizations of incident light.

2.4 Different Polarizations

So far all results have assumed p polarization as depicted in figure 2.1. This section will investigate the slight changes that arise in the case of s polarization. Because the math has already been worked out in detail once, and because the case of s polarization is somewhat simpler, the math shown here will be kept brief.

From the geometry depicted in figure 2-2, it is clear that the wave vector ex-

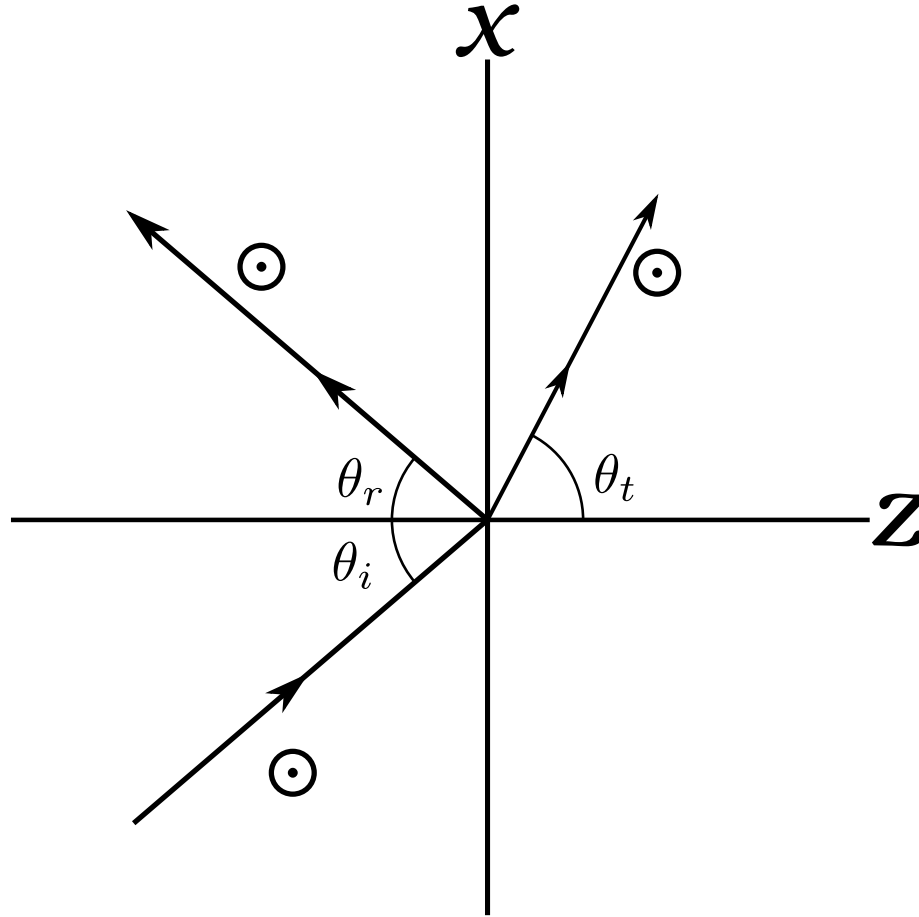


Figure 2-2: Geometry for s polarization. Large arrows represent propagation direction of plane waves. Arrow tip indicators show that all electric fields point in the +y direction.

expressions remain unchanged from equation 2.1. As a consequence, the spacial and temporal dependence in the electric fields do not change, and the law of reflection and Snell's law can be derived exactly as they were in section 2-1. Even the complex transmission angle, θ_t , is defined exactly as it was for p polarization. However, the electric field expressions do change and it turns out they simplify considerably. The electric fields simplify so much that the boundary condition for the displacement field used to derive the reflection coefficient in the case of p polarized light is satisfied trivially and does not provide useful information to calculate the reflection coefficient. Instead the boundary conditions for tangential and normal magnetic fields suffice.

As before, these conditions are also well established [6]. The system of equations analogous to equations 2.9 is similar.

$$E_{0i} \cos \theta_i - E_{0r} \cos \theta_i = E_{0t} \cos \tilde{\theta}_t \quad (2.16a)$$

$$E_{0i} + E_{0r} = \frac{n_1}{\tilde{n}_2} E_{0t} \quad (2.16b)$$

After solving the system, it is clear that the amplitude reflection coefficient for s polarized light, denoted \tilde{r}_s , is similar to but simpler than the coefficient for p polarized light. In terms of refractive indices and known angles, \tilde{r}_s can be expressed as,

$$\tilde{r}_s = \frac{n_1 \cos \theta_i - \tilde{n}_2 \cos \tilde{\theta}_t}{n_1 \cos \theta_i + \tilde{n}_2 \cos \tilde{\theta}_t} \quad (2.17)$$

whereas, in terms of permittivities and wave vector components, it can be written

$$\tilde{r}_s = \frac{k_{iz} - \tilde{k}_{tz}}{k_{iz} + \tilde{k}_{tz}} \quad (2.18)$$

Because the wave vectors have already been calculated in the previous section, and discussion of their sign has been saved for the next section, the intensity reflection coefficient can be stated immediately.

$$R_s = \frac{(k_{iz} - k'_{tz})^2 + k''_{tz}{}^2}{(k_{iz} + k'_{tz})^2 + k''_{tz}{}^2} \quad (2.19)$$

2.5 Different Quadrants for \tilde{k}_{tz}

The final task is to determine an expression for the transmitted wave number's z-component, \tilde{k}_{tz} . Using $\tilde{k}_{tz}^2 = \tilde{k}_t^2 - k_{tx}^2$ with $\tilde{k}_t^2 = (\epsilon'_2 + i\epsilon''_2)k_0^2 = (n_2^2 - \gamma^2 - 2in_2\gamma)k_0^2$,

\tilde{k}_{tz}^2 can be expressed in terms of either refractive indices or permittivities as,

$$\tilde{k}_{tz}^2 = k_0^2 (n_2^2 - \gamma^2 - n_1^2 \sin^2 \theta_i - 2in_2\gamma) = k_0^2 (\epsilon'_2 - \epsilon_1 \sin^2 \theta_i + i\epsilon''_2) \quad (2.20)$$

and \tilde{k}_{tz} can then be evaluated as,

$$\tilde{k}_{tz} = \pm k_0 \sqrt{n_2^2 - \gamma^2 - n_1^2 \sin^2 \theta_i - 2in_2\gamma} = \pm k_0 \sqrt{\epsilon'_2 - \epsilon_1 \sin^2 \theta_i + i\epsilon''_2} \quad (2.21)$$

The fact that \tilde{k}_{tz} is given by the square root of equation 2.20 means that it will have two possible conjugate values. This indeterminacy of \tilde{k}_{tz} has led authors to draw very different conclusions about ATIR and the incidence angles at which it can exist. To proceed toward finding the solutions for \tilde{k}_{tz} 's real and imaginary parts, the real and imaginary parts of equation 2.20 can be equated to give a system of coupled equations in k'_{tz} , and k''_{tz} .

$$k'^2_{tz} - k''^2_{tz} = k_0^2(n_2^2 - \gamma^2 - n_1^2 \sin^2 \theta_i) \quad (2.22a)$$

$$k'_{tz}k''_{tz} = -k_0^2 n_2 \gamma \quad (2.22b)$$

Solving 2.22b first for k'_{tz} , then for k''_{tz} , and substituting into 2.22a yields the following decoupled quadratics for k'^2_{tz} and k''^2_{tz} .

$$\left. \begin{aligned} k'^4_{tz} - [n_2^2 - \gamma^2 - n_1^2 \sin^2 \theta_i]k_0^2 k'^2_{tz} - k_0^4 n_2^2 \gamma^2 &= 0 \\ k''^4_{tz} + [n_2^2 - \gamma^2 - n_1^2 \sin^2 \theta_i]k_0^2 k''^2_{tz} - k_0^4 n_2^2 \gamma^2 &= 0 \end{aligned} \right\} \quad (2.23)$$

or

$$\left. \begin{aligned} k'^4_{tz} - [\epsilon'_2 - \epsilon_1 \sin^2 \theta_i]k_0^2 k'^2_{tz} - \frac{k_0^4 \epsilon''^2_2}{4} &= 0 \\ k''^4_{tz} + [\epsilon'_2 - \epsilon_1 \sin^2 \theta_i]k_0^2 k''^2_{tz} - \frac{k_0^4 \epsilon''^2_2}{4} &= 0 \end{aligned} \right\} \quad (2.23')$$

Because the gain parameters, γ and ϵ''_2 , appear squared in equations 2.23, both k'_{tz}

and k''_{tz} have the same magnitudes regardless of whether the second medium is gainy or lossy. The fact that equations (2.23) are quartic suggests that there may be four solutions for \tilde{k}'_{tz} and another four solutions for \tilde{k}''_{tz} given as follows.

$$k'_{tz} = \pm \frac{k_0}{\sqrt{2}} \sqrt{\epsilon'_2 - \epsilon_1 \sin^2 \theta_i \pm \sqrt{(\epsilon'_2 - \epsilon_1 \sin^2 \theta_i)^2 + \epsilon_2''^2}} \quad (2.24)$$

$$k''_{tz} = \pm \frac{k_0}{\sqrt{2}} \sqrt{\epsilon_1 \sin^2 \theta_i - \epsilon'_2 \pm \sqrt{(\epsilon'_2 - \epsilon_1 \sin^2 \theta_i)^2 + \epsilon_2''^2}} \quad (2.25)$$

However, k'_{tz} and k''_{tz} were defined in equation 2.12 to be strictly real. Consequently, the signs preceding the inner square roots above must both be positive, reducing the preceding equations to,

$$k'_{tz} = \pm \frac{k_0}{\sqrt{2}} \sqrt{\epsilon'_2 - \epsilon_1 \sin^2 \theta_i + \sqrt{(\epsilon'_2 - \epsilon_1 \sin^2 \theta_i)^2 + \epsilon_2''^2}} \quad (2.26)$$

$$k''_{tz} = \pm \frac{k_0}{\sqrt{2}} \sqrt{\epsilon_1 \sin^2 \theta_i - \epsilon'_2 + \sqrt{(\epsilon'_2 - \epsilon_1 \sin^2 \theta_i)^2 + \epsilon_2''^2}} \quad (2.27)$$

Equation 2.22b dictates that the real and imaginary parts of \tilde{k}_{tz} must have opposite signs for $\gamma > 0$ (gainy second medium) and the same sign for $\gamma < 0$ (lossy second medium). This restricts the complex quantity \tilde{k}_{tz} to lie in either the second or fourth quadrant in the case of a gainy second medium, and the first or third quadrant in the case of a lossy second medium. So \tilde{k}_{tz} has indeed been reduced to two possible solutions as was expected initially from equation 2.21.

The ultimate question that needs to be answered is which of the two mathematically valid solutions for p_{tz} is physically correct in each of the four cases where the incidence is below or above the critical angle on a lossy or gainy medium. To answer the question, we must first observe that the z -dependent phase factor in the plane wave is given by

$$e^{i\tilde{k}_{tz}z} = e^{ik'_{tz}z - k''_{tz}z} \quad (2.28)$$

The physically correct solution is defined by the following three conditions.

1. A wave's amplitude must grow in the propagation direction in a gainy medium and decay in a lossy medium.
2. The solution corresponding to a non-evanescent wave behind the boundary, when $\theta_i < \theta_c$ and $|k''_{tz}| \ll k'_{tz}$, must propagate away from the boundary in either lossy or gainy media.
3. The solution corresponding to an evanescent wave behind the boundary, when $\theta_i > \theta_c$ and $|k''_{tz}| \gg k'_{tz}$, must decay away from the boundary in either lossy or gainy media.

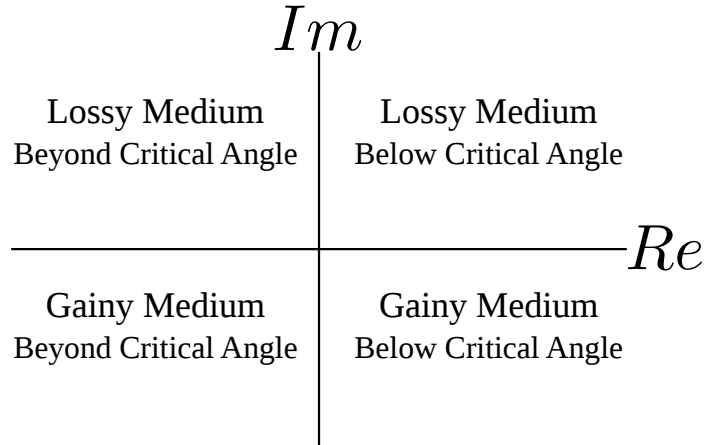


Figure 2-3: \tilde{k}_{tz}^2 will lie in different quadrants depending on the gain factor's sign and the incident angle.

Figure 2-3 summarizes the conditions under which \tilde{k}_{tz}^2 lies in various quadrants of the complex plane. In the next sections, each of the four cases will be discussed, and the proper root, \tilde{k}_{tz} , will be chosen.

2.5.1 Lossy Medium Below Critical Angle

We will begin by investigating the case of a lossy second medium because the solution is not debated, and it will serve as an introduction to the case of a gainy

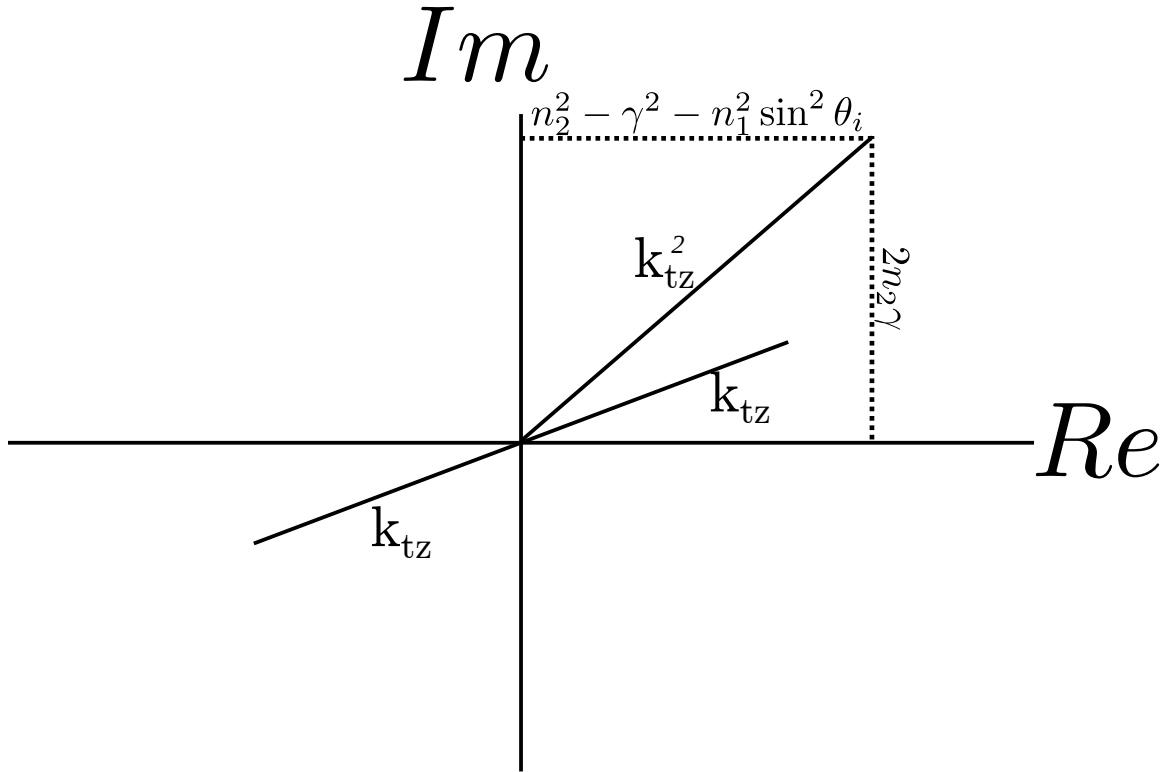


Figure 2-4: For lossy second medium and incidence below the critical angle, \tilde{k}_{tz}^2 lies in the first quadrant. Its conjugate roots lie in the first and third quadrants.

second medium, which is widely debated. When the incident angle is below the critical angle, the quantity \tilde{k}_{tz}^2 lies in the first quadrant as illustrated in figure 2-4. The two mathematical solutions for \tilde{k}_{tz} in this case lie in the first and third quadrants.

The imaginary part of \tilde{k}_{tz}^2 is determined only by properties of the media, whereas the real part depends on incident angle as well. As the incident angle increases from normal incidence up to the critical angle, the location of \tilde{k}_{tz}^2 in the complex plane traces a horizontal line to the left such that it lies along the positive imaginary axis at the critical angle.

Choosing the correct physical solution for \tilde{k}_{tz} , will dictate whether the reflected wave is amplified or attenuated. The two possible reflectivities, calculated from equation 2.15, are plotted in figure 2-5. The z -dependent phase factor from equation 2.28

Plane wave reflectivity options, R , for incidence on a lossy medium, $\theta_i < \theta_c$

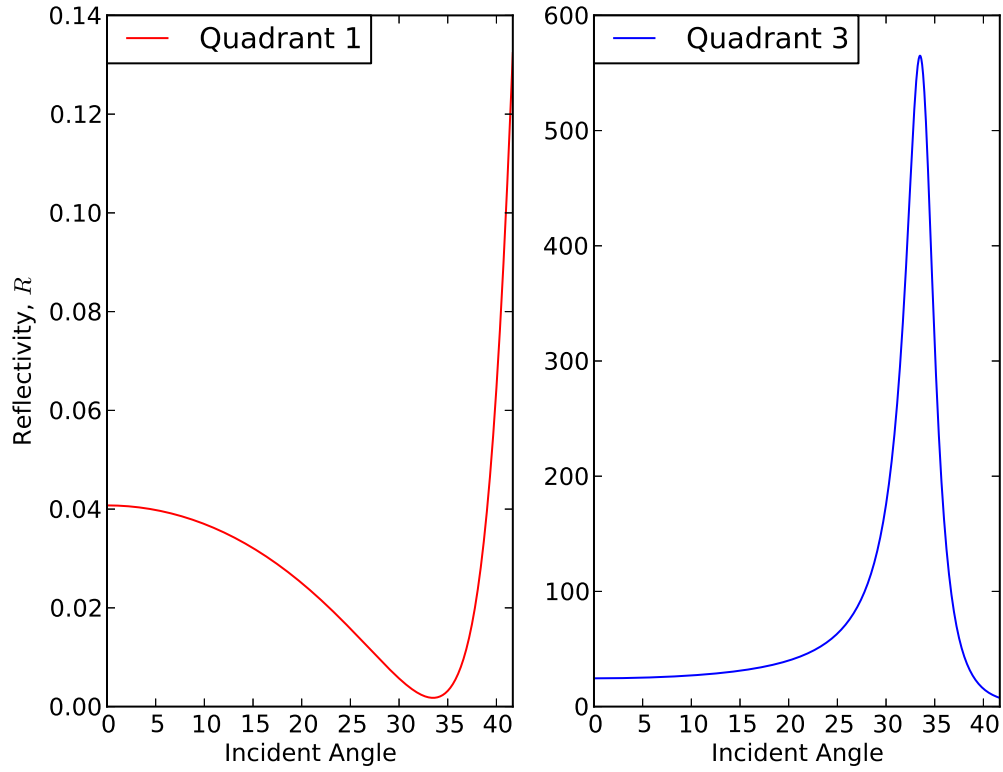


Figure 2-5: The two curves depict the possible reflectivities, R from the lossy medium at incidence below the critical angle. Only the curve below unity is physically correct.

can be written with all signs explicitly shown as,

$$e^{\tilde{k}_{tz}z} = \begin{cases} e^{i|k'_{tz}|z - |k''_{tz}|z} & 1^{st} \text{ Quadrant} \\ e^{-i|k'_{tz}|z + |k''_{tz}|z} & 3^{rd} \text{ Quadrant} \end{cases} \quad (2.29)$$

so that the three previously stated conditions can be more easily used to determine the quadrant of \tilde{k}_{tz} . Condition one requires that the wave's amplitude decay in its propagation direction. The Q1 solution is right-traveling, and decaying as z increases, so it satisfies the first condition. The Q3 solution is left-traveling and decaying as z decreases, so it also satisfies the first condition. Because this wave is incident below

the critical angle, it is not evanescent and condition two requires that it travel away from the boundary, in the $+z$ direction. The sign in front of $|k''_{tz}|$ above indicates that only the Q1 solution meets this requirement, so we can conclude that it is the physically correct solution.

2.5.2 Lossy Medium Beyond Critical Angle

When the incidence angle increases beyond the critical angle, \tilde{k}_{tz}^2 moves from the first quadrant into the second following the same horizontal line that it did previously. The two conjugate roots remain in the first and third quadrants.

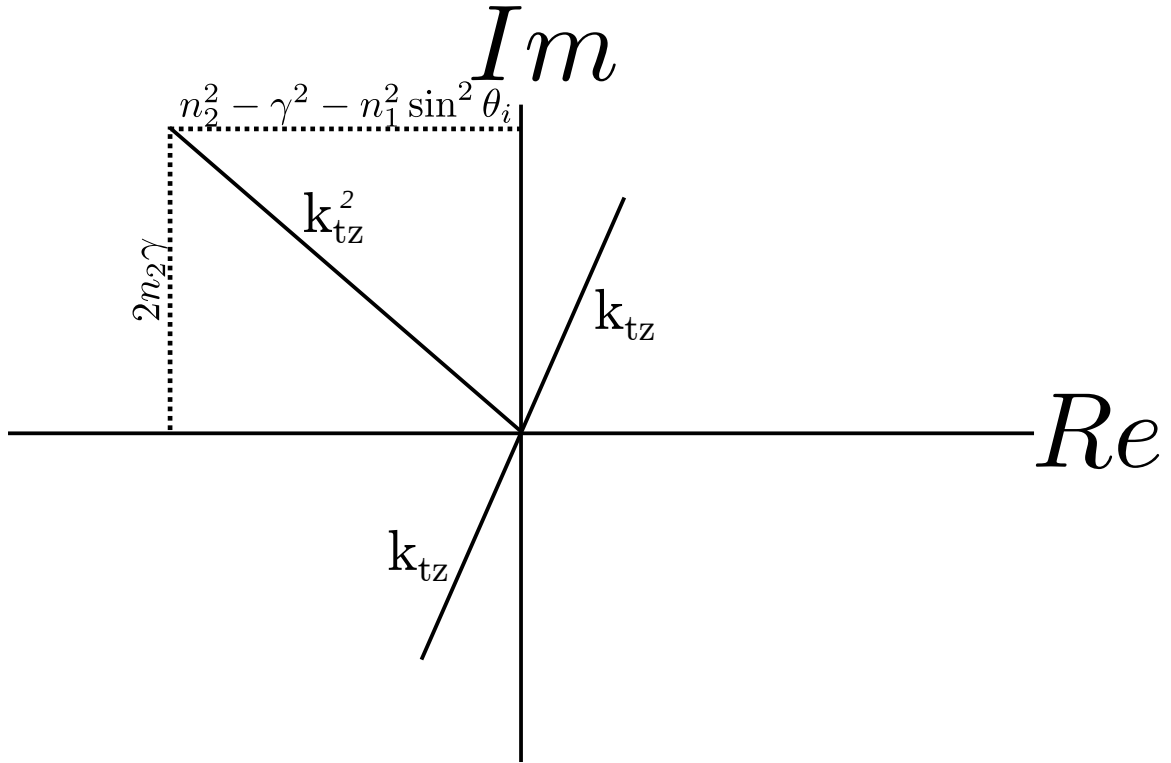


Figure 2-6: For lossy second medium and incident angle beyond the critical angle, \tilde{k}_{tz}^2 lies in the second quadrant. Its conjugate roots remain in the first and third quadrants.

As before, the z -dependent phase factor from equation 2.28, written with all signs explicitly shown, is

$$e^{i\tilde{k}_{tz}z} = \begin{cases} e^{i|k'_{tz}|z - |k''_{tz}|z} & 1^{st} \text{ Quadrant} \\ e^{-i|k'_{tz}|z + |k''_{tz}|z} & 3^{rd} \text{ Quadrant} \end{cases} \quad (2.30)$$

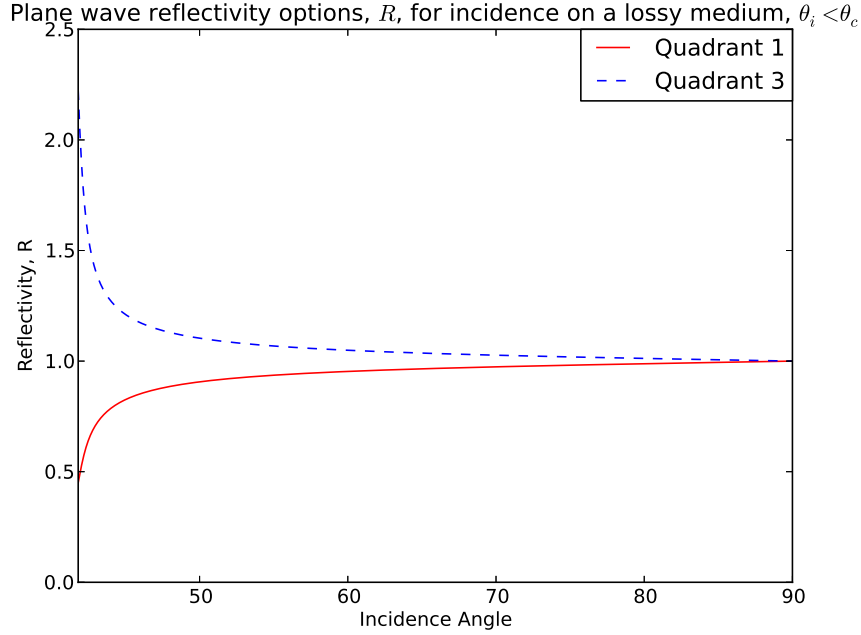


Figure 2-7: The two curves depict the possible reflectivities, R from the lossy medium at incidence beyond the critical angle. Only the curve below unity is physically correct.

The two resulting reflectivity options from equation 2.15 are shown in figure 2-7.

The first of the conditions on page 23 is satisfied by both solutions as it was for incidence below the critical angle. However, the second condition, which determined the physically correct solution for incidence below the critical angle, no longer applies because the wave in medium two has become evanescent. Rather the third condition requires that the evanescent wave decays away from the boundary. The positive sign in front of $|k''_{tz}|$ makes it clear that the Q3 solution will grow exponentially in medium two, leaving the Q1 solution, which decays away from the boundary, as the physically correct solution.

In addition to meeting the requirements outlined for the physically correct solu-

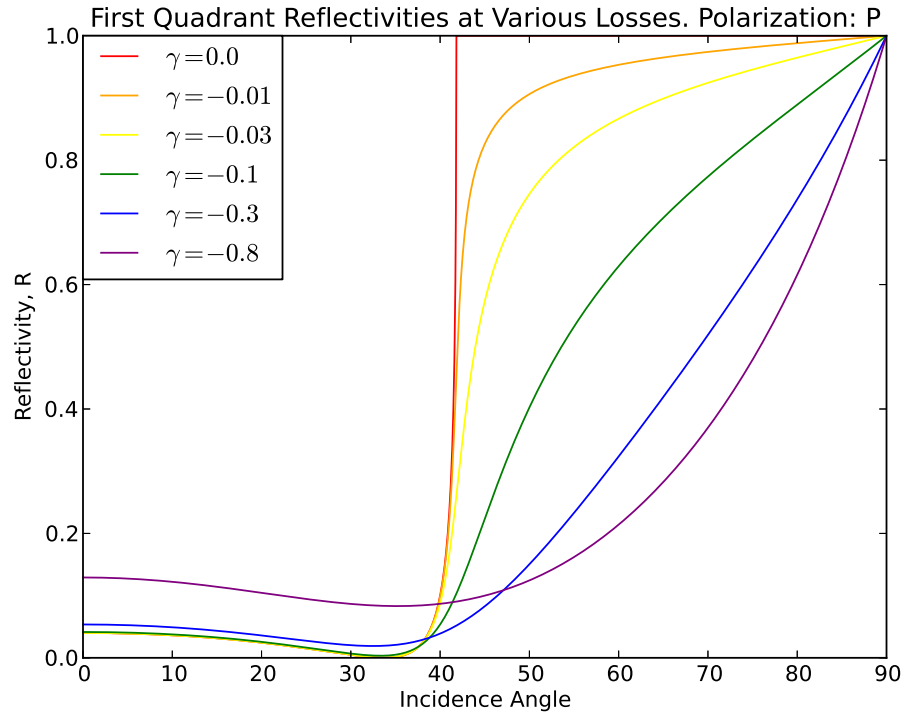


Figure 2-8: First quadrant reflectivity, R , from lossy second medium plotted for various values of loss. Notice that for any nonzero loss, the reflectivity does not reach zero at the Brewster angle.

tion, the Q1 solution provides attenuated TIR both below and beyond the critical angle. The Q3 solution would provide amplified TIR violating energy conservation. This is further assurance that satisfying the three conditions has led to the physically correct solution.

Having chosen the Q1 solution for a lossy second medium both below and beyond the critical angle, the complete reflectivity curves are plotted for various values of loss in figure 2-8

2.5.3 Gainy Medium Below Critical Angle

Having resolved the lossy second medium case, we can now discuss the debated case of a gainy second medium. When a wave is incident below the critical angle on

a gainy second medium, \tilde{k}_{tz}^2 lies in the fourth quadrant and the conjugate roots, have now shifted from the first and third quadrants to the second and fourth as depicted in figure 2-9. As in the case of a lossy second medium, as the incident angle increases from the normal to the critical angle, \tilde{k}_{tz}^2 moves horizontally to the left until it is resting on the negative imaginary axis.

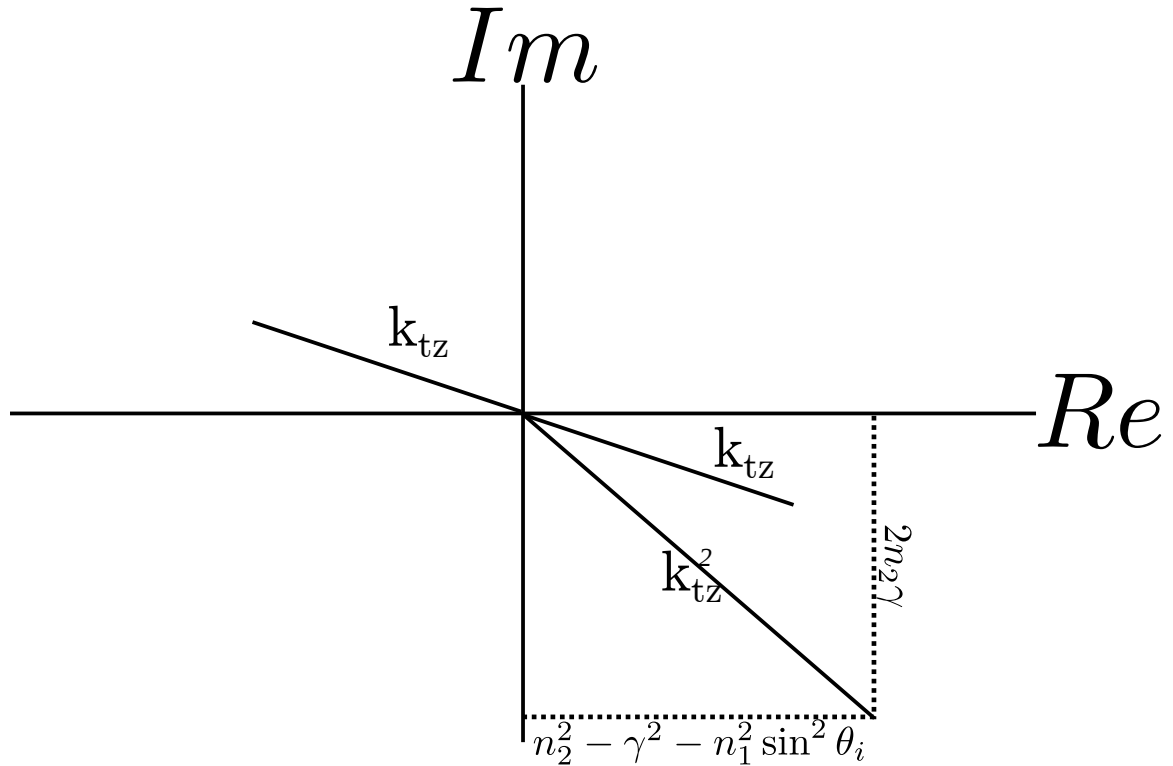


Figure 2-9: For gainy second medium and incident angle below the critical angle, \tilde{k}_{tz}^2 lies in the fourth quadrant. Its conjugate roots lie in the fourth and second quadrants.

The plane wave reflectivity options from equation 2.15, resulting from the new quadrants for \tilde{k}_{tz} , are plotted in figure 2-10. The curves are the same as figure 2-5 in the lossy case, although the quadrant for \tilde{k}_{tz} is importantly not. This illustrates the important general rule that reflectivities will be the same for conjugate loss and gain media if the corresponding quadrants for \tilde{k}_{tz} are chosen. This is a consequence of the fact that the gain factor, γ appears squared in equations 2.23.

Plane wave reflectivity options, R , for incidence on a gainy medium, $\theta_i < \theta_c$

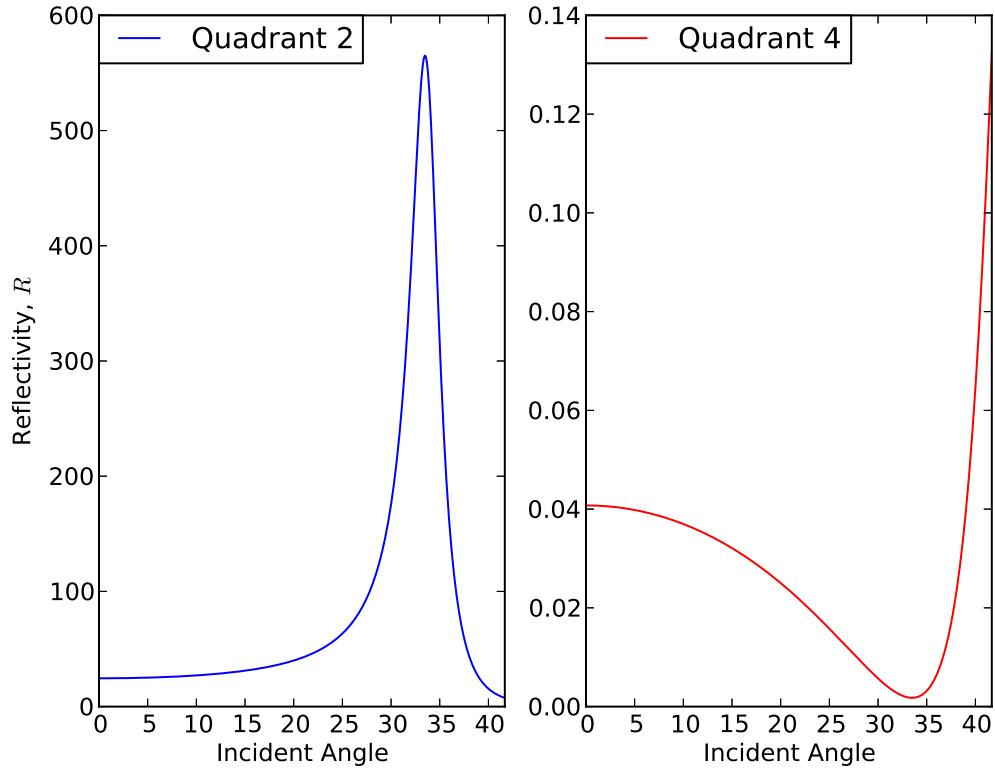


Figure 2-10: The two curves depict the possible reflectivities, R from a gainy medium at incidence below the critical angle. Only the curve below unity is physically correct.

In the literature, authors are not unanimously agreed on the physically correct quadrant for \tilde{k}_{tz} in this case as they are in the case of lossy media. Most authors, including [13] [4] [14] and [11], have chosen the fourth quadrant solution leading to attenuated reflection below the critical angle, whereas the authors of [2] have chosen the second quadrant solution leading to Amplified TIR. As usual, the physically correct solution can be discerned by writing the z dependent phase factor from equation 2.28. Unlike the lossy case, equation, 2.22b requires that real and imaginary parts of \tilde{k}_{tz} have the same sign in the gainy case.

$$e^{i\tilde{k}_{tz}} = \begin{cases} e^{-i|k'_{tz}|z - |k''_{tz}|z} & 2^{nd} \text{ Quadrant} \\ e^{i|k'_{tz}|z + |k''_{tz}|z} & 4^{th} \text{ Quadrant} \end{cases} \quad (2.31)$$

The second quadrant solution is left-traveling and its amplitude grows as it travels, whereas the fourth quadrant solution is right-traveling and also growing as it travels. So the first condition is again satisfied for both solutions. Because the transmitted wave is non-evanescent, condition two requires that it propagate away from the boundary. The fourth quadrant solution satisfies this condition, but the left-traveling second quadrant solution does not. The fourth quadrant solution is the only physically correct solution. It is not useful to plot the solution at various gains because the curves would be identical to those of the conjugate loss medium in figure 2-8 below the critical angle.

Callary and Carniglia, the authors of [2], were led to support the existence of the second quadrant solution in this case because they began by analyzing the three layer problem – two transparent media with a gainy slab between them. They concluded that waves corresponding to both the Q2 and the Q4 solutions would exist in the slab because reflection from the third layer, would convert one solution to the other. Using their three layer results in the limit as $d \rightarrow \infty$, they concluded that amplified reflection was possible from a single surface. Their mistake derived from the fact that in the true two layer problem, there is no back surface from which a wave can reflect. The authors viewed their result as a theoretical explanation of the gain of 1000 observed by the Kogan et. al. experiment [10]. But while their theory did explain the high gain observed experimentally, the experimental result was a consequence of the three layer configuration that they treated rather than a strictly single surface phenomenon.

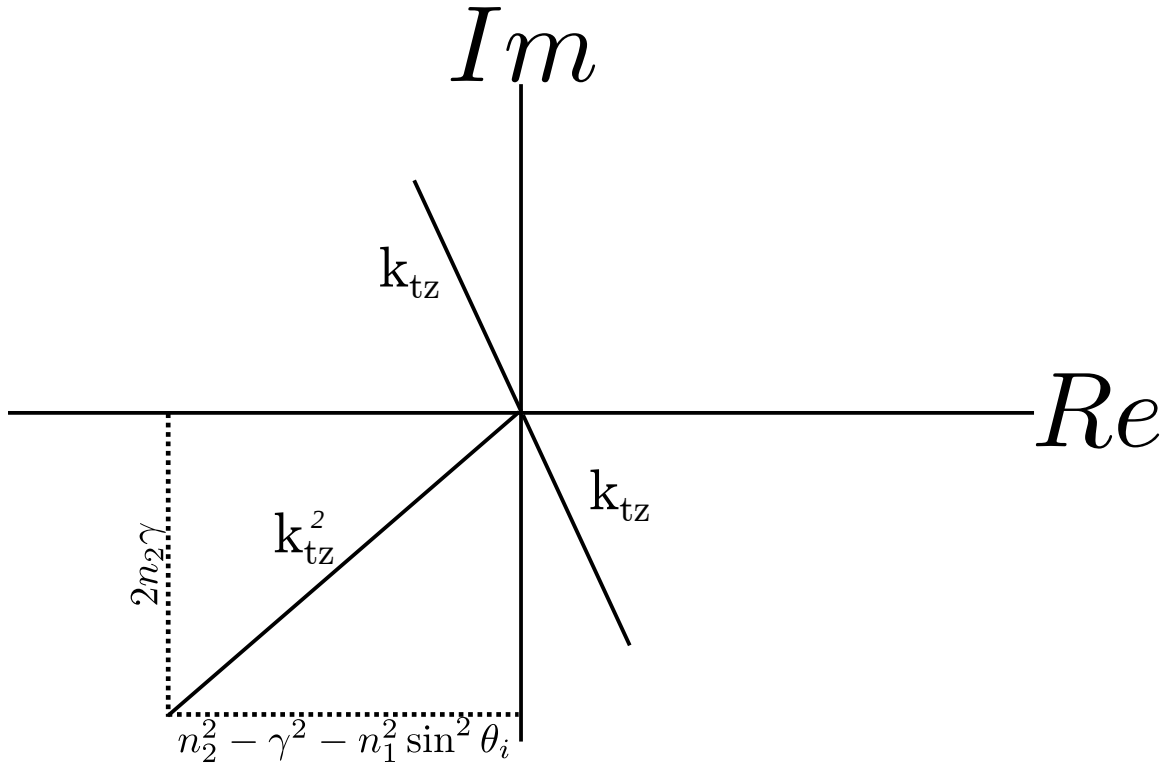


Figure 2-11: For gainy second medium and incident angle beyond the critical angle, \tilde{k}_{tz}^2 lies in the third quadrant. It's conjugate roots remain in the fourth and second quadrants.

2.5.4 Gainy Medium Beyond Critical Angle

The final case that remains to be analyzed is incidence beyond the critical angle on a gainy second medium. Under these conditions, the incidence angle has become great enough that \tilde{k}_{tz}^2 has moved into the third quadrant. The possible conjugate roots remain in the second and fourth quadrants as they were for incidence below the critical angle.

In this case, the z dependent phase factor can be written as it was in the previous section.

$$e^{i\tilde{k}_{tz}z} = \begin{cases} e^{-i|k'_{tz}|z - |k''_{tz}|z} & 2^{nd} \text{ Quadrant} \\ e^{i|k'_{tz}|z + |k''_{tz}|z} & 4^{th} \text{ Quadrant} \end{cases} \quad (2.32)$$

It has already been stated that the first of the three conditions is satisfied by both solutions. Beyond the critical angle, the transmitted wave is evanescent, so the third condition requires that its amplitude decay away from the boundary. The negative sign in front $|k''_{tz}|z$ in the Q2 solution indicates that it meets this requirement, while the positive sign in front of $|k''_{tz}|z$ in the Q4 solution indicates an exponentially growing wave. This leaves Q2 as the only physically correct solution.

The two possible options have not been plotted beyond the critical angle because they are, again, identical to those in figure 2-7. The reflectivities are shown at all incident angles in figures 2-12 and 2-13 for various values of the gain factor, γ .

The Mansuripur, the authors of [11], argue that the second quadrant solution is “unlikely” on the grounds that as the incidence angle increases from just below the critical angle to just above it, \tilde{k}_{tz} jumps discontinuously from the fourth quadrant to the second quadrant. Because no such discontinuous jump exists in the case of classical TIR from transparent media, where \tilde{k}_{tz} is strictly real below the critical angle, zero at that angle, and strictly imaginary above it, the authors argue that no discontinuity need exist for gainy media either. Their argument is incorrect for two reasons. First, the case of TIR from transparent media is a special case of that which is presently being treated in which \tilde{k}_{tz} does not jump discontinuously at the critical angle only because the Q4 and Q2 solutions are degenerate, both being equal to zero. Second, in the case of lossy or gainy media, a discontinuity in \tilde{k}_{tz} *must* exist under some conditions. Choosing the Q2 solution eliminates the discontinuity at the critical angle, but creates one when a wave is incident beyond the critical angle and the second media is tuned from gainy to lossy. This is illustrated in the summarizing diagram in the next section.

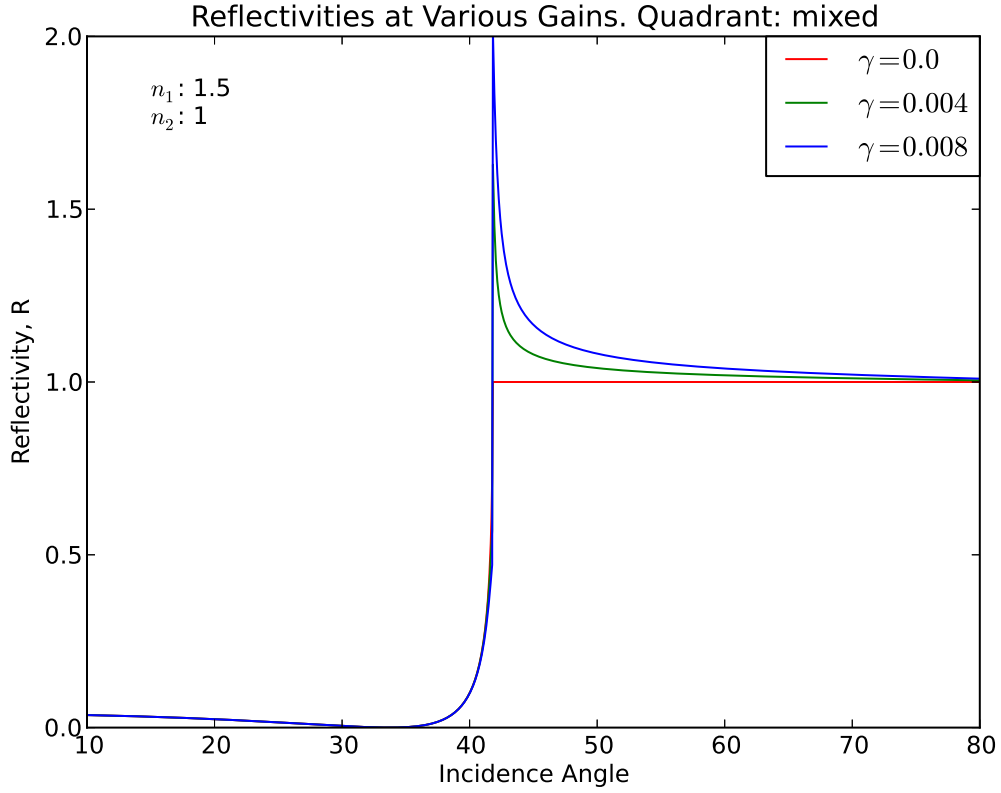


Figure 2-12: Surface reflectivity R for plane waves on a gainy second medium. Reflectivity is discontinuous at the critical angle where k_{tz} switches from the fourth quadrant to the second.

2.5.5 Conclusions and Experimental Evidence

Based on the arguments in the previous sections, I choose the first quadrant solution for \tilde{k}_{tz} at all incidence angles if the second medium is lossy. But, if the second medium is gainy, the fourth quadrant solution is chosen only below the critical angle, while the second quadrant solution takes over beyond θ_c , resulting in ATIR, beyond the critical angle.

Figure 2-14 summarizes this information in the context of a hypothetical experiment. Imagine a lab setup such that both incident angle and second media gain/loss were controlled by knobs. The system is initially configured such that \tilde{k}_{tz}^2 lies at

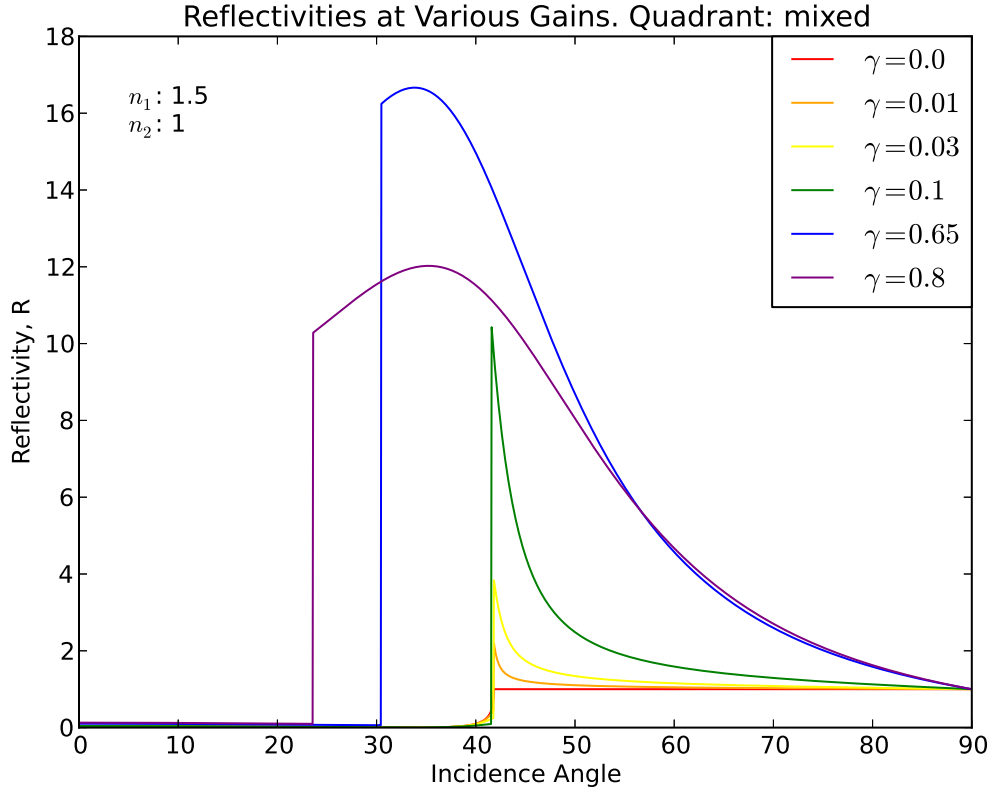


Figure 2-13: Surface reflectivity R for plane waves on a gainy second medium. With higher gain present it becomes clear that there is a maximum reflectivity for given n_1 and n_2 . It also becomes clear that the critical angle moves closer to the normal as gain increases.

point A' corresponding to an incident wave striking a lossy medium at a glancing angle with \tilde{k}_{tz} at point A on the inner locus. As the incidence angle is tuned from glancing, through the critical angle, down to near normal incidence, \tilde{k}_{tz}^2 and \tilde{k}_{tz} advance to points B' and B . Maintaining near normal incidence, γ is increased, from lossy, through transparent, to gainy, and \tilde{k}_{tz}^2 and \tilde{k}_{tz} reach points C' and C . Callary and Carniglia disagree about the location of point C , and claim that \tilde{k}_{tz} experiences a discontinuous jump to the second quadrant as the medium becomes transparent which is indicated with a dotted line. With the second medium now gainy, the incidence angle is again tuned through the critical angle, back to glancing incidence, and

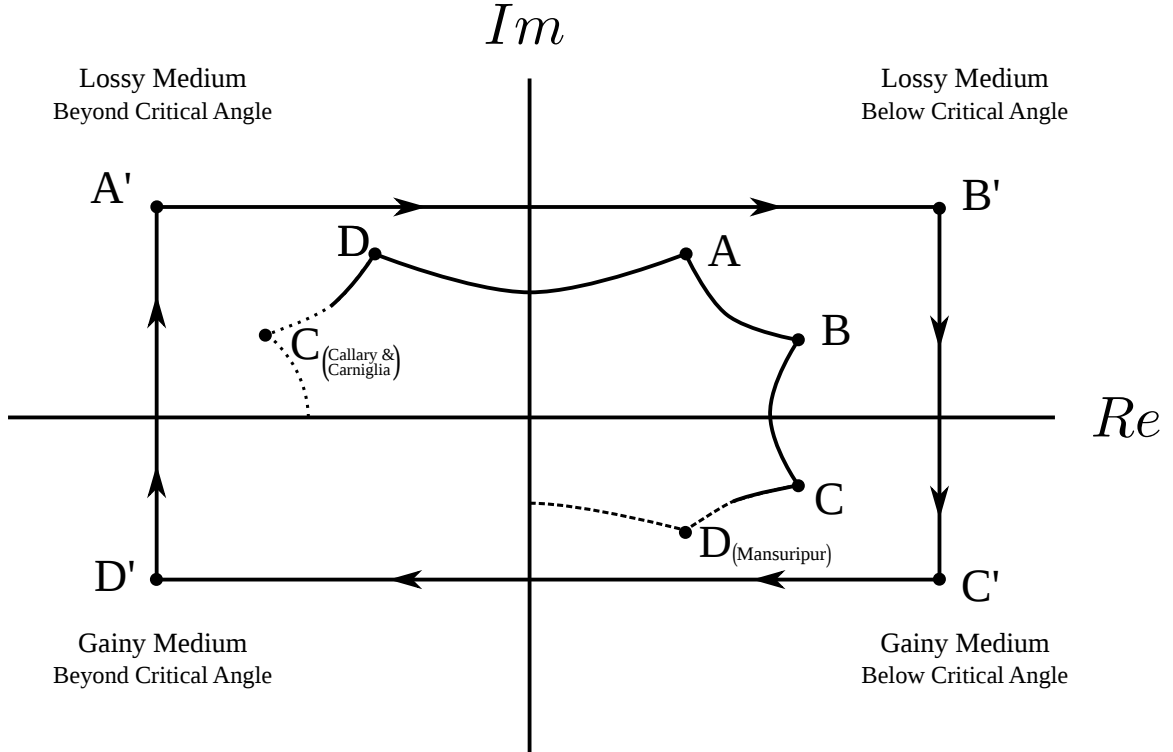


Figure 2-14: Detailed quadrant summary of \tilde{k}_{tz}^2 (rectangle) and \tilde{k}_{tz} (solid inner locus) under various conditions. Dotted line indicates Callary and Carniglia's alternate locus. Dashed line indicates Mansuripur's alternate locus. No theory suggests a locus in the third quadrant – it would mean ATIR from lossy media.

the loci advance to points D and D' . This time it is the Mansuripur who disagree about the location of point D , and their alternate locus is indicated with a dashed line. Finally γ is tuned back to lossy and the loci return to points A and A' .

The theory and quadrant choices outlined in this section are able to explain all experimental data to date. The Koester fiber experiment deals with multiple reflections, beyond the critical angle, so even a small gain on each reflection could account for his results. The gain of 25 observed by Kogan et.al. is feasible from single surface ATIR, and can also be explained according to the three layer theory of Callary and Carniglia. The gain of 1000 observed by Kogan et. al. could only be explained by single surface ATIR if the second quadrant were chosen for incidence angles below the

critical angle, which I have argued to be physically incorrect. Rather their exceedingly high gain must be explained in terms of a third layer according to the theory of Callary and Carniglia.

Chapter 3

Case of a Finite Beam

As was mentioned in section 1.1, a real world light beam is never a true plane wave. In order to make meaningful predictions about experimental results, I will now treat the case of a finite width beam. The exact profile of the beam is not critically important; any finite beam will do. It is customary in the literature [9][14] to treat a Gaussian profile, and I will do likewise. The Gaussian profile is very convenient because it closely approximates the true profile of a laser beam produced from a cavity in its $TEM_{0,0}$ mode.

Until now, all results have been analytical and exact. The formulas derived for plane wave reflection in equations 2.11, 2.15, 2.18, and 2.19 were in closed form without approximation. In this chapter approximations and numerical work will be necessary to treat the additional mathematical complexity of finite beams and Fourier analyses. I will discuss and use methods to minimize error throughout the chapter, but ultimately approximate results will still be quite useful as the main intention is to gain understanding of how finite beams differ from plane waves, not to exactly solve for the reflectivity of a particular finite beam profile.

3.1 Derivation and Properties of a Gaussian Beam

The derivation of a Gaussian beam is well established, and found in many text books (eg. [1]), so it will be outlined here only briefly. Attention rather will be placed on the properties and parameters of the beam.

It is convenient to define a set of primed coordinates that are the beam's natural coordinates, such that the beam travels primarily along the z' direction. Starting from the Helmholtz wave equation,

$$\nabla^2 \tilde{A} + k^2 \tilde{A} = 0, \tag{3.1}$$

if the cross-sectional profile of the beam changes slowly as the beam propagates, the so-called paraxial approximation, $\partial A/\partial z \ll kA$, can be made to reduce equation 3.1 to its paraxial form.

$$\frac{\partial \tilde{A}}{\partial x^2} + \frac{\partial \tilde{A}}{\partial y^2} + 2ik \frac{\partial \tilde{A}}{\partial z} = 0 \tag{3.2}$$

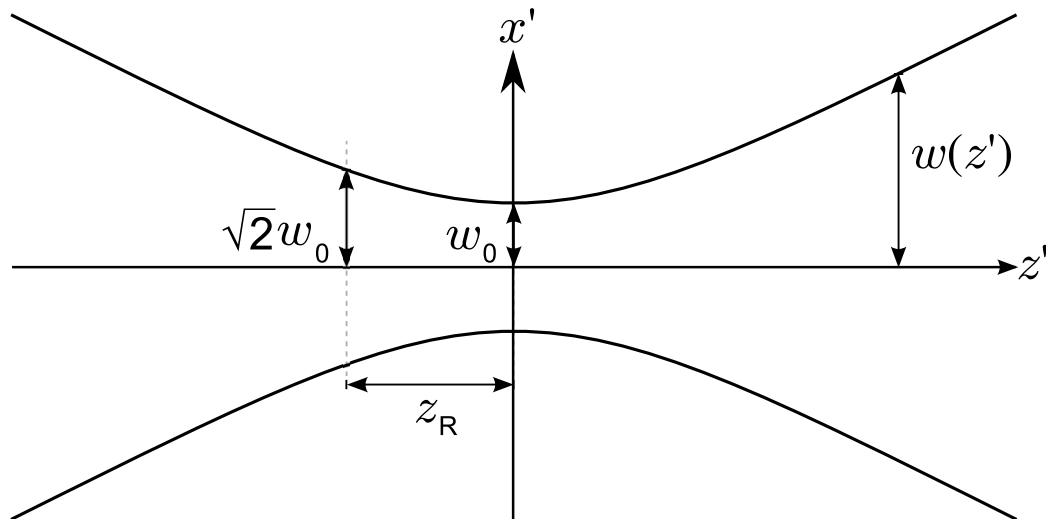


Figure 3-1: Width of a Gaussian beam near its waist (at $z' = 0$). Rayleigh range and beam width are shown.

In the case of a wave traveling in the z' direction the complex electric field amplitude is given by $E(\mathbf{x}') = \tilde{A}(\mathbf{x}')e^{ikz'}$, and the field of a Gaussian beam satisfying the paraxial wave equation, has the general form,

$$E(\mathbf{x}') = E_0 \frac{w_0}{w^2(z')} \exp \left[-\frac{x'^2 + y'^2}{w^2(z')} + ik \left(z' - \frac{x'^2 + y'^2}{2R(z')} \right) + i\xi(z') \right] \quad (3.3)$$

where it remains to define the quantities $w(z')$, $R(z')$, and $\xi(z')$.

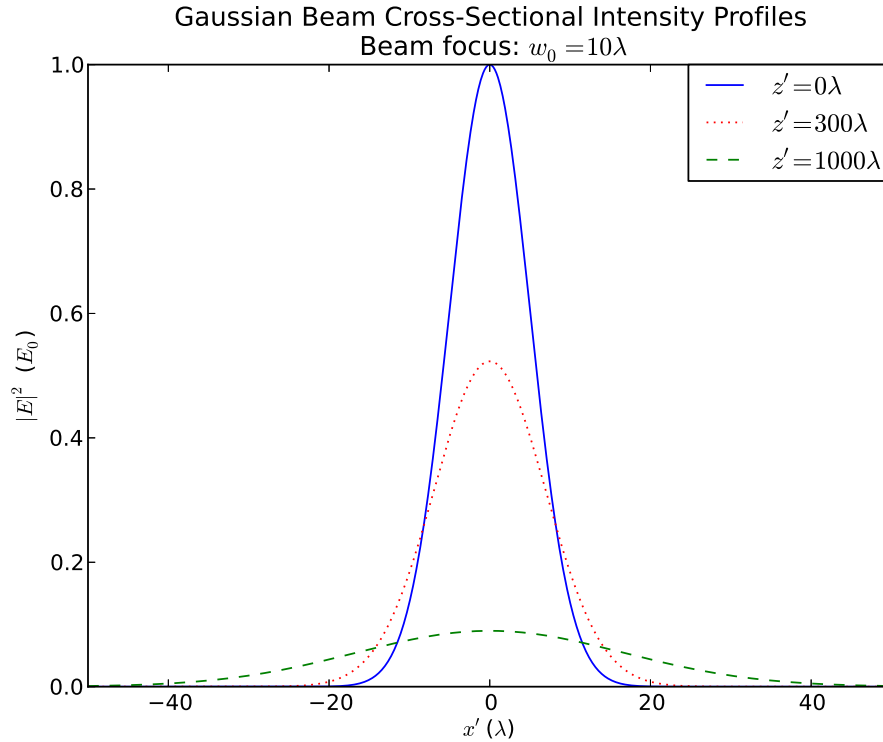


Figure 3-2: Cross-section intensity profiles of a Gaussian beam at various longitudinal positions. The cross section is always Gaussian, its radius is not constant.

The cross-sectional intensity profile of the beam described in equation 3.3 is Gaussian at any point as seen in figure 3-2, but its radius is not constant. The beam reaches a narrowest point, $w_0 = w(0)$, known as its waist, at $z' = 0$. In the paraxial approximation, plane waves must travel primarily in the beam's propagation direction. In order to keep this approximation valid, the beam cannot be focused more narrowly

than about one wavelength, $w_0 > \lambda$. The beam width at any point other than $z = 0$ is defined as the radius at which the electric field amplitude falls to $(1/e)$ -th of its value on the central axis. This is illustrated in figure 3-1 and given functionally as follows.

$$w(z) = w_0 \sqrt{1 + \left(\frac{z'}{z_R}\right)^2} \quad (3.4)$$

The additional longitudinal distance, z_R , is referred to as the Rayleigh range and defined as the propagation distance at which the beam waist has increased by a factor of $\sqrt{2}$, also illustrated in figure 3-1.

$$z_R = \frac{\pi w_0^2}{\lambda} = \frac{w_0^2 k}{2} \quad (3.5)$$

The surfaces of constant phase in the beam make spherical wave fronts which are described in terms of their radius, $R(z')$. It is important that R is not measured from the primed coordinate origin to a phase front, but rather from the center of the phase front's arc to the phase front itself as illustrated in figure 3-3.

$$R(z') = z' \left(1 + \left(\frac{z_R}{z'}\right)^2\right) \quad (3.6)$$

Because the Gaussian beam is composed of many plane waves traveling in slightly different directions, its phase behaves in an unusual way. First of all the phase is not constant in x' (or y') for fixed z' . Second, the beam goes through an additional phase shift of π radians near its waist which is described by the Gouy phase, $\xi(z')$, given as follows.

$$\xi(z') = \arctan(z'/z_R) \quad (3.7)$$

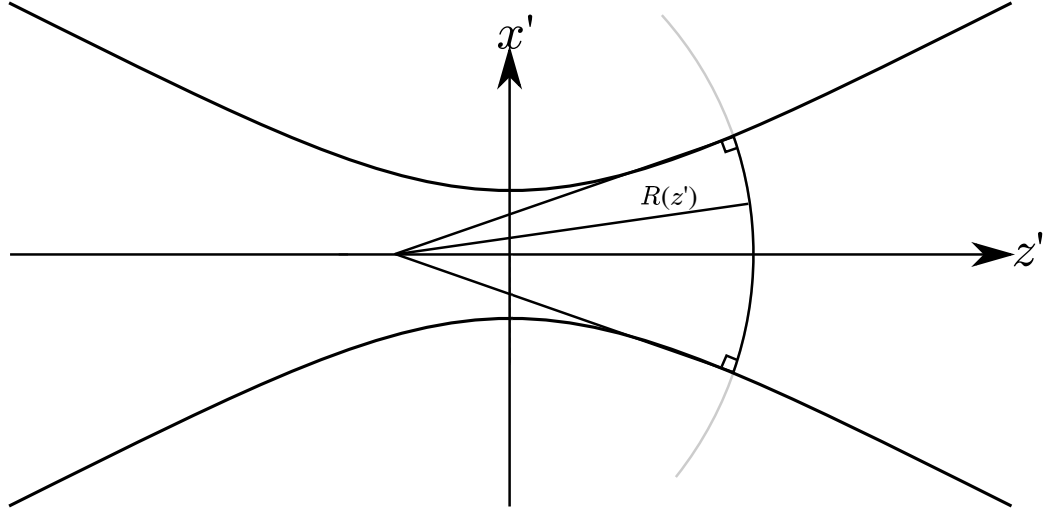


Figure 3-3: Gaussian beam near its waist. Phase front curvature, $R(z')$ is shown. Note: $R(z')$ is not measured from the coordinate origin.

3.1.1 Coordinate Systems for Reflection

The Gaussian beam is expressed most naturally and easily in its natural coordinate system. It will therefore be necessary to relate the natural coordinates of the incident and reflected beams to the surface coordinates used until now. The Gaussian beam's central ray strikes the surface at angle ϕ . This nomenclature is chosen to distinguish the special central ray from the constituent plane waves in the rest of this chapter. Figure 3-4 shows the three coordinate systems as well as the beam widths for the incident and reflected beams. Note that all of the coordinate systems are simply related to one another by rotations through the incidence (and reflection) angle, ϕ . In particular, transformations between the three sets of axes can be easily accomplished using the following relations.

$$z' = z \cos \phi + x \sin \phi \quad x' = x \cos \phi - z \sin \phi \quad (3.8a)$$

$$x = x' \cos \phi + z' \sin \phi \quad z = z' \cos \phi - x' \sin \phi \quad (3.8b)$$

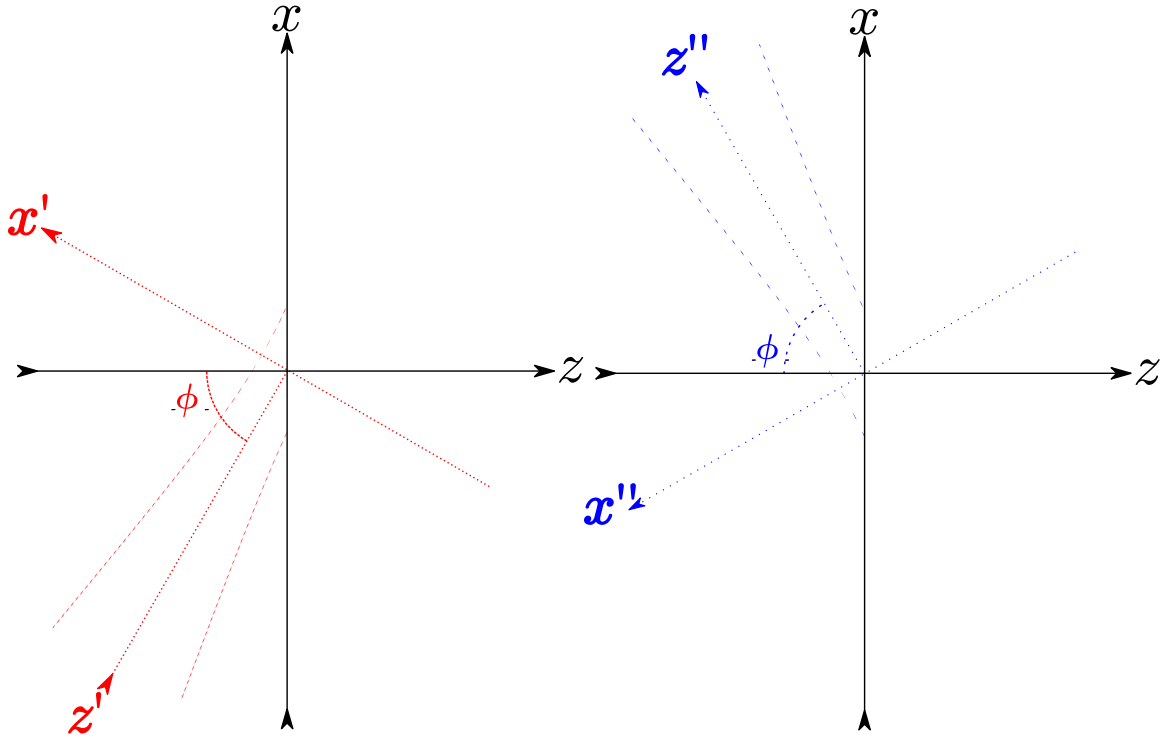


Figure 3-4: Natural axes of incident (left, primed) and reflected (right, doubly primed) Gaussian Beams. z' and z'' axes are the primary propagation directions. Curved lines show the beam width along the propagation directions.

$$z'' = -z \cos \phi_i + x \sin \phi \quad x'' = -x \cos \phi - z \sin \phi \quad (3.9a)$$

$$x = -x'' \cos \phi + z'' \sin \phi \quad z = -z'' \cos \phi - x'' \sin \phi \quad (3.9b)$$

3.2 Fourier Analysis of Finite Beams

Having thoroughly treated the case of incident plane waves in the previous chapter, the behavior of a finite width beam can be understood by expressing it as a superposition of constituent plane waves. This is accomplished through use of the Fourier transform, which expresses the wave's electric field as a function of parallel wave number components, k_{ix} , or plane wave incident angles, θ_i . That is, the Fourier

transform describes how much of the finite beam's intensity is present in each of the constituent plane waves that travel in slightly different directions. For this treatment, it will not be necessary to consider the direction perpendicular to the incidence plane (the y -direction), so this coordinate will be suppressed for the remainder of the chapter, and electric fields will be written as functions of x and z only.

To begin, the beam must be written in terms of its natural coordinate system as in equation 3.3. The transform can be computed at any location, and in any of the three coordinate systems to provide information about the intensity contained in each plane wave. But in order to get the correct reflection from the surface, the phase information, and consequently the complex electric field, must be known at the surface. Therefore, it is necessary to use the coordinate system transformation equations 3.8 to write the incident electric field in terms of unprimed coordinates.

$$E_i(x', z')|_{z=0}(x) = E_i(x \cos \phi, x \sin \phi) \quad (3.10)$$

In particular,

$$E_i(x) \approx \frac{E_0}{\sqrt{1 + \frac{x^2 \sin^2 \phi}{z_R^2}}} \exp \left[\frac{-x^2 \cos^2 \phi}{w_0^2 \left(1 + \frac{x^2 \sin^2 \phi}{z_R^2}\right)} + ikx \sin \phi - ik \frac{x \cos^2 \phi}{2 \sin \phi \left(1 + \frac{z_R^2}{x^2 \sin^2 \phi}\right)} + i \arctan \left(\frac{x \sin \phi}{z_R} \right) \right] \quad (3.11)$$

Using the surface expression for the incident electric field, the Fourier transform can be taken to give the necessary amplitude and phase information at the surface.

$$E_i(k_x) = \int_{-\infty}^{\infty} E_i(x) e^{-ik_x x} dx \quad (3.12)$$

With the incident beam now decomposed into constituent plane waves, each plane wave can be reflected from the surface according to the theory of chapter 2 using

equation 2.11 for p -polarization or equation 2.18 for s -polarization.

$$E_r(k_x) = E_i(k_x)\tilde{r}(k_x). \quad (3.13)$$

The reflected beam can then be reconstructed by integrating over all reflected plane waves. Note that the kernel of the inverse transform can include an additional z dependent term used to propagate the beam away from the reflection surface, with the z -component wave numbers calculated as $k_z = \sqrt{k^2 - k_x^2}$. Or, alternatively, the transform can be written in terms of plane wave incidence angles rather than wave vector components.

$$E_r(x, z) = \int_{-\infty}^{\infty} E_r(k_x) e^{i(k_x x + k_z z)} dk_x = \int_{-\pi/2}^{\pi/2} E_r(\theta_i) e^{ik(\sin \theta_i x + \cos \theta_i z)} d\theta_i. \quad (3.14)$$

If it is necessary or convenient, the reflected beam can be transformed into its own natural coordinates, the doubly primed coordinates, by use of equations 3.9.

In the next two sections, the steps outlined here will be performed in detail both analytically and numerically and the differences between the two methods will be discussed.

3.2.1 Analytical Method with Approximations

The expression for the incident field at the reflecting surface given in equation 3.11 is as exact as the expression in equation 3.3. The only approximation so far is that it is a solution to the paraxial form of the Helmholtz equation. Consequently, the only limitation to keep in mind is that the spot size must be larger than about one wavelength. However, the expression for $E_i(x)$ on the surface is too complex to be transformed analytically according to the integral in equation 3.12, so two more approximations will be necessary.

The first issue to consider is the arctan term that arises from the Gouy phase. Using a small angle approximation, this term can be simplified to $\arctan\left(\frac{z'}{z_R}\right) \approx \frac{z'}{z_R}$, which leaves off all terms in the Taylor series expansion of $\mathcal{O}(z'^3)$ and higher. The approximation is quite good when the beam strikes the surface near normal incidence, but becomes arbitrarily bad as the incidence angle, ϕ , increases. Alternately a large angle approximation allows the simplification $\arctan\left(\frac{z'}{z_R}\right) \approx \frac{\pi}{2} - \frac{z_R}{z'}$, which leaves off all terms of $\mathcal{O}(z'^{-3})$. Again, this approximation is good at glancing incidence, but arbitrarily bad near normal incidence. Because we want to treat the full domain of incidence angles, and because the true arctan never exceeds $\pm\pi/2$, it is a good approximation to neglect the Gouy phase entirely.

$$\arctan\left(\frac{z'}{z_R}\right) \approx 0 \tag{3.15}$$

This approximation simplifies the integration considerably, and introduces less error over the full range of incidence than either of the previously mentioned alternatives. If a higher degree of accuracy is desired for a particular incidence angle, the Taylor series expansion can be performed around that point. As we will see, a first order expansion never increases the difficulty of analytical integration.

Removing the Gouy phase from the exponential is helpful, but further simplification will be necessary to perform the Fourier transform analytically. The second approximation is to assume that the beam focus remains wide enough that for all values of x on the surface, within the beam's width,

$$x^2/z_R^2 \ll 1, \tag{3.16}$$

which reduces the beam width to the constant w_0 , and allows the 1 term in the curvature radius to be neglected leaving $R(z') \approx z_R^2/z'$. The approximation can then be applied again, eliminating this term entirely because it is small compared to the

ikz' term.

These two approximations allow the form for the incident field at the surface to simplify from the form in equation 3.11 to the following,

$$E_i(x) \approx E_0 \exp \left[\frac{-x^2 \cos^2 \phi}{w_0^2} + ikx \sin \phi \right]. \quad (3.17)$$

Before continuing directly to the integration, it is valuable to investigate how accurate this approximation is, and to compare the approximated expression to the unapproximated form in equation 3.11. This comparison is made graphically in figure 3-5. The two points to observe are that the approximation is better for wider beam waists, and that phase is affected more than intensity. Even the narrowest beam shown ($w_0 = 1\lambda$) does not exhibit significant error. As long as the beam width required for the paraxial approximation is met, the approximations made in this section will be justified.

The integral in equation 3.12 can now be written in the following form.

$$E_i(k_x) \approx E_0 \int_{-\infty}^{\infty} \exp [Ax^2 + Bx] dx \quad (3.18)$$

$$A = \frac{-\cos^2 \phi}{w_0^2} \quad (3.19)$$

$$B = ik (\sin \phi - \sin \theta_i) \quad (3.20)$$

which is known to integrate to

$$E_i(k_x) \approx \frac{\sqrt{\pi}}{\sqrt{-A}} \exp[-B^2/4A], \quad (3.21)$$

thus providing an analytical, but approximate, Fourier decomposition into constituent plane waves. It is convenient to write the expression in terms of the plane waves' incident angles, θ_i .

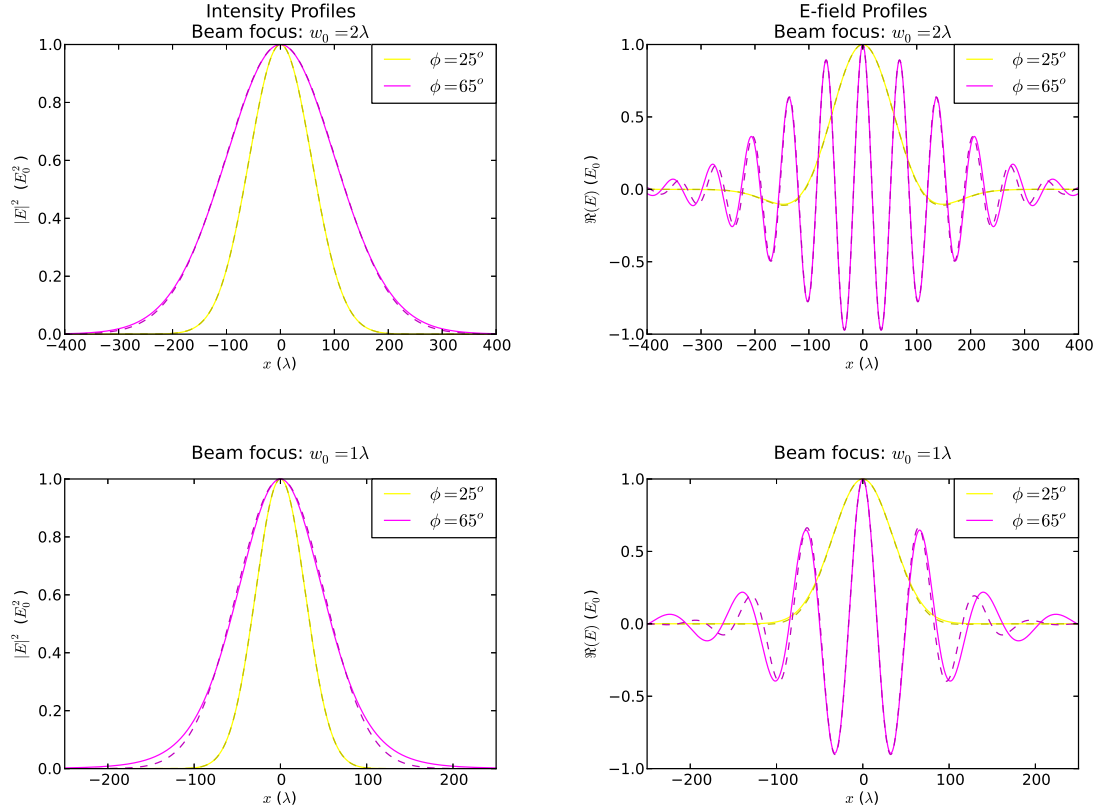


Figure 3-5: Comparison of exact (solid curves) and approximated (dashed curves) Gaussian beams at the reflecting surface. Left graphs show intensity vs surface position. Right graphs show electric field vs surface position. Narrow beams are affected more than wide beams, and phase is affected more than intensity.

$$E_i(\theta_i) \approx \frac{E_0 w_0 \sqrt{\pi}}{\cos \phi} \exp \left[-\frac{k^2 w_0^2 (\sin \phi - \sin \theta_i)^2}{4 \cos^2 \phi} \right] \quad (3.22)$$

The reflected beam is calculated at any point in the plane of incidence by performing the integral in equation 3.14. Despite the approximations made, the inverse transform integral must be computed numerically.

3.2.2 Numerical Method – the Discrete Fourier Transform

As an alternative to making approximations to the incident field in equation 3.11, the Fourier integral can be performed numerically by converting the integral 3.12 into a discrete sum, and sampling the integrand accordingly.

The mathematics of the Discrete Fourier Transform (DFT) are well established [3], and follow naturally from the continuous transform outlined in the previous section. The discretization of the Fourier integral from equation 3.12 is as follows.

$$\tilde{F}_m(k_m) = \sum_{n=0}^{N-1} E_n(x_n) e^{-ik_m x_n} \quad (3.23)$$

Naturally, calculating the transform discretely requires that the function being transformed, $E_i(x)$ presently, be sampled at a finite number of points. For this reason, the DFT is most often used when the input signal is recorded experimentally. The spacial domain function is sampled at N evenly-spaced locations numbered $x_0, \dots, x_n, \dots, x_{N-1}$, with the sampled values labeled, $E_0, \dots, E_n, \dots, E_{N-1}$. The sum in equation 3.23 is computed N times to provide N complex transformed Amplitudes, F_m , at corresponding x -component wave numbers k_m . The sampling process is illustrated in figure 3-6. The actual functions depicted in the figure are not of critical importance, they are simply used to illustrate the sampling process and related quantities. The figure depicts the following quantities (with example units to assist in physical intuition), where κ_s is the spacial sampling frequency.

$$x_n = nL = \text{the } n^{\text{th}} \text{ sample position (meters)}$$

$$k_m = mK = \text{the } m^{\text{th}} \text{ wave number sample (radians per meter)}$$

$$L = 1/\kappa_s = \text{the spacial sampling interval (meters)}$$

$$K = 2\pi\kappa_s/N = \text{wave number sampling interval (rad/meter)}$$

Using the relations above, the wave number samples are given by,

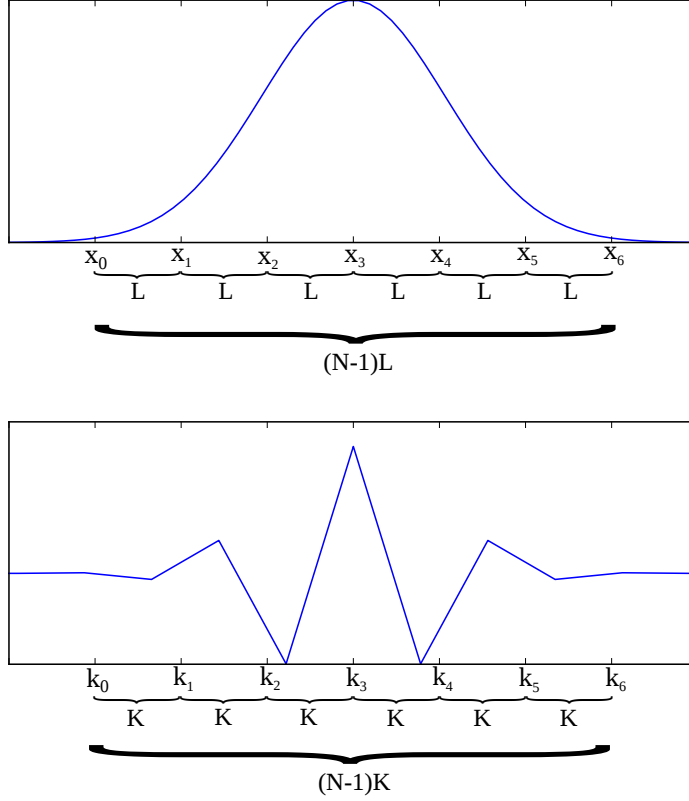


Figure 3-6: Schematic diagram illustrating sampling intervals and nomenclature in x and k domains for a DFT. In this illustration, $N = 7$.

$$k_m = \frac{2\pi m}{LN} \quad (3.24)$$

Of course, as N increases, so does the accuracy of the DFT. But, in a physical problem, unlike a purely mathematical problem, the value of the x -component wave numbers, k_m , cannot exceed the wave number supported by the physical medium in which the Gaussian beam is traveling, $-k_i \leq k_m \leq k_i$. In order to achieve this, the spacial sampling interval, L , must remain above the minimum value given by $L \geq \lambda/2$, and N cannot increase without a corresponding increase in the size of the sampling domain. If care is taken to ensure these requirements, the DFT is effective, and can be computed very quickly with an algorithm known as the fast Fourier transform (FFT) [3].

From this point, the process follows similarly to that outlined in the previous section. The reflection coefficients, $\tilde{r}(k_x)$ are sampled at the same k_x values at which incident amplitudes are calculated, and an inverse transform is taken. It is important that the FFT algorithm can only be used to compute the inverse transform if it is also performed on the reflecting surface. If the reflected beam is to be propagated away from the surface to a detector by changing the transform kernel, the sum must be computed by other means which results in a significant increase in computation time. This is the primary weakness of the DFT method.

3.2.3 Comparison of the Two Methods

Each of the two methods above has its advantages and disadvantages, and by understanding them, the most appropriate method can be executed for each application.

The analytical method has the advantage that it can be used to propagate a reflected beam away from the boundary with no significant increase in computation time or difficulty, but it has the disadvantage that approximations must be made to the expression for the incident electric field. The most notable effect of these approximations is that the analytical Fourier transform is a strictly real function. The analytical method's second disadvantage is that the inverse transform must still be computed numerically, so numerical error will be introduced despite the approximations made.

The DFT method has the advantage of speed. The FFT algorithm makes performing the sums very fast. It also has the advantage that the only errors introduced are numerical, and consequently they can be made arbitrarily small by choosing high enough sample numbers. But its disadvantage is that the inverse transform can only be evaluated at the reflecting surface.

Figure 3-7 shows the Fourier transform of a Gaussian beam incident at two different angles both numerically and analytically. Note that the analytical transform is strictly real, and a continuous function, whereas the discrete transform is sampled

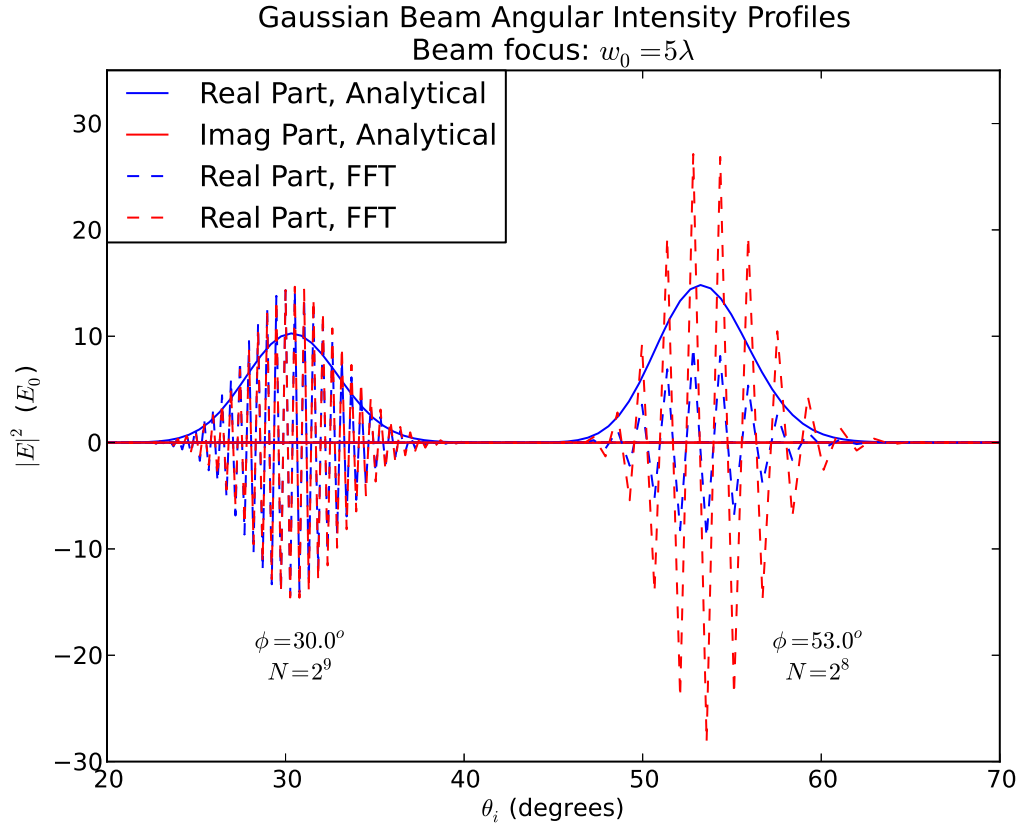
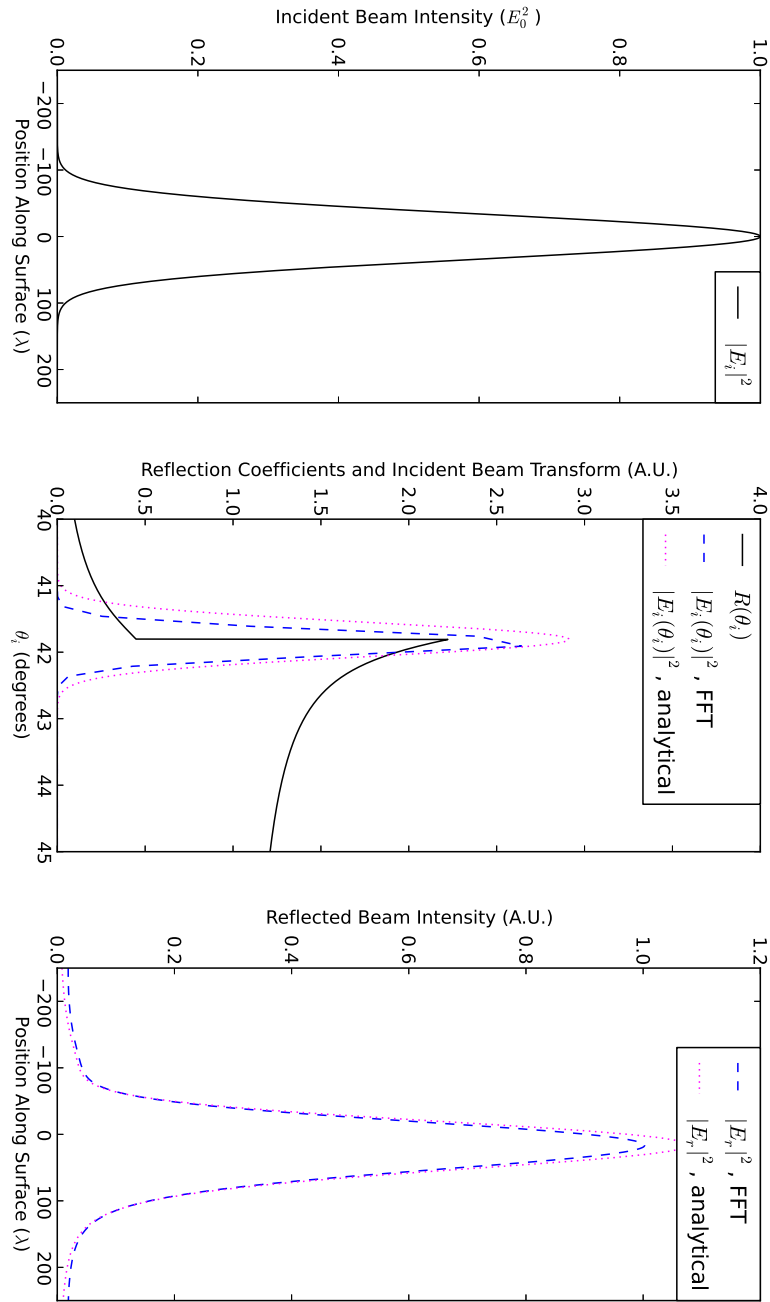


Figure 3-7: Comparison of analytical and FFT Fourier transform methods for two different incident angles and Sample sizes. Significantly, the analytical transform is strictly real, leaving the curve for its imaginary part lying on axis. Also illustrated is the effect of sample number, N in the FFT.

at a finite set of points and is, in general, complex. The discrete sampling interval is different for the two incident angles, illustrating the effects of course versus fine sampling.

The differences between the analytical and FFT methods can also be seen in the surface reflection profiles they generate. Figure 3-8 shows the surface profile of an incident beam in the left pane. The same incident profile plotted against the θ_i domain using both the analytical and FFT methods is shown in the center pane. The FFT spectrum shows signs of sampling effect, which could easily be eliminated



Comparison of incident and reflected intensity profiles of Gaussian beams in p polarization
Incidence at critical angle, $\phi = \theta_c$

Figure 3-8: Comparison of incident and reflected intensity profile using both analytic and discrete methods. Slight pulse reshaping occurs when incidence is near the critical angle.

by increasing the sample size. The spectra are superimposed over the reflection coefficients to depict which of the constituent plane waves are below and beyond the critical angle. The third pane shows the surface profile of the reflected beam calculated by both methods. Note that the curve for the analytical method has been scaled so the curves do not lie directly on top of one another.

3.3 Finite Beam Reflection Profiles Under Many Conditions

Having analyzed the case of a finite beam and discussed two methods to perform the relevant calculations, it will be insightful to plot the reflectivities of finite beams under various conditions. All of the graphs in this section will be computed by means of the FFT for speed considerations and because calculations at points of the reflecting surface will not need to be considered.

First I will consider a series of Gaussian beams of different waist sizes incident on a gainy reflecting surface. Figure 3-9 shows p polarized beams of widths between two and twelve wavelengths. The intensity vs incidence angle plot shows that the widely focused beam follows most closely to the discontinuous graph for a plane wave, while the narrowly focused beam exhibits reflectivity that increases slowly and smoothly. Figure 3-10 shows the same information for s polarization.

Figure 3-11 holds the beam width fixed at $w_0 = 10\lambda$ and plots the reflectivities vs incidence angle for several values of gain, γ . Although increasing γ to arbitrarily high values does not increase reflectivity without bound as shown in 2-13, increasing gain within the realm of physically realizable values does.

Lastly, in figure 3-12, reflectivity is plotted against γ for a few incident angles. At a fixed incidence angle and spot size ($w_0 = 10\lambda$), it is clear that increasing gain only increases reflectivity to an extent before it falls off again. Additionally, the

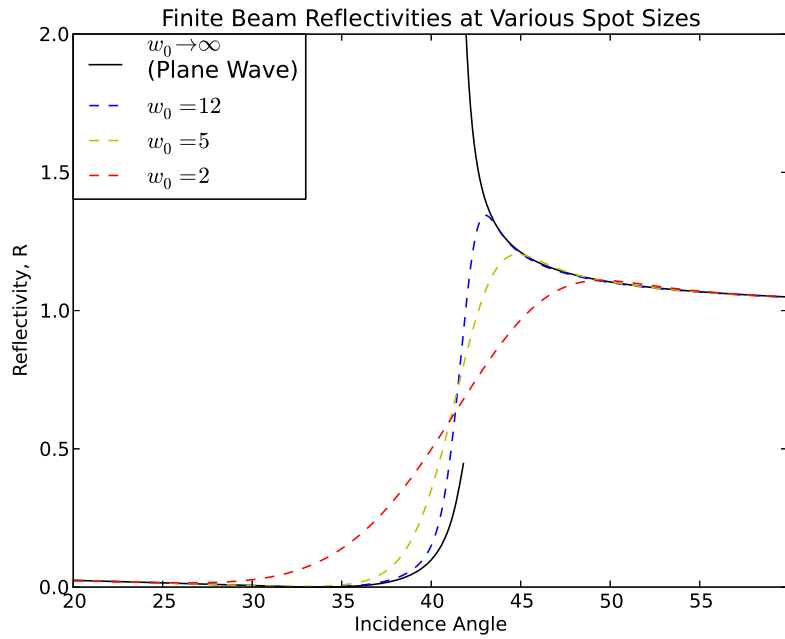


Figure 3-9: Finite beam reflectivities at spot sizes of 2, 5, and 12 wavelengths (p polarization). The widely focused beam follows most closely to the plane wave reflectivity.

beam incident at $\phi = 30^\circ$ illustrates the fact that the critical angle changes as γ changes. The beam is incident below the critical angle at first, but as gain increases through $\gamma = 0.6$, the critical angle falls below the incidence angle and the beam begins experiencing ATIR.

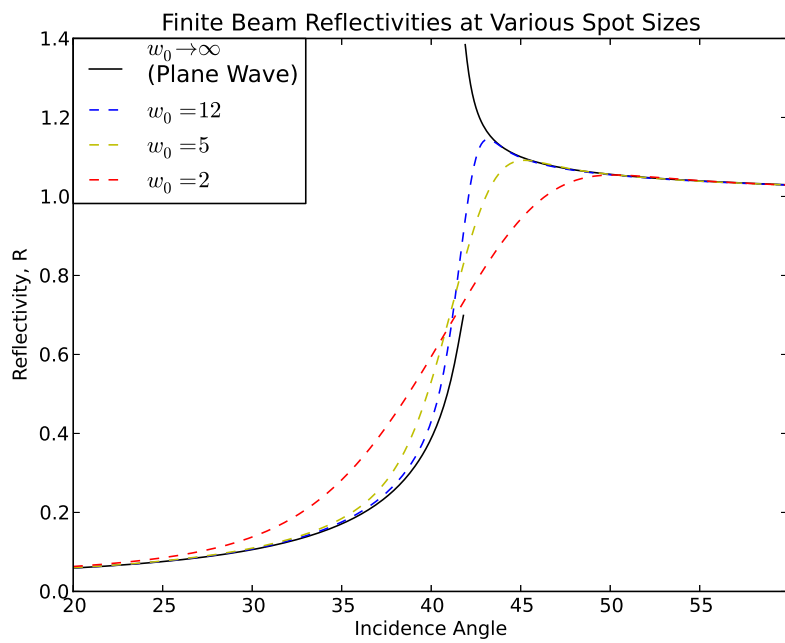


Figure 3-10: Finite beam reflectivities at spot sizes of 2, 5, and 12 wavelengths (s polarization).

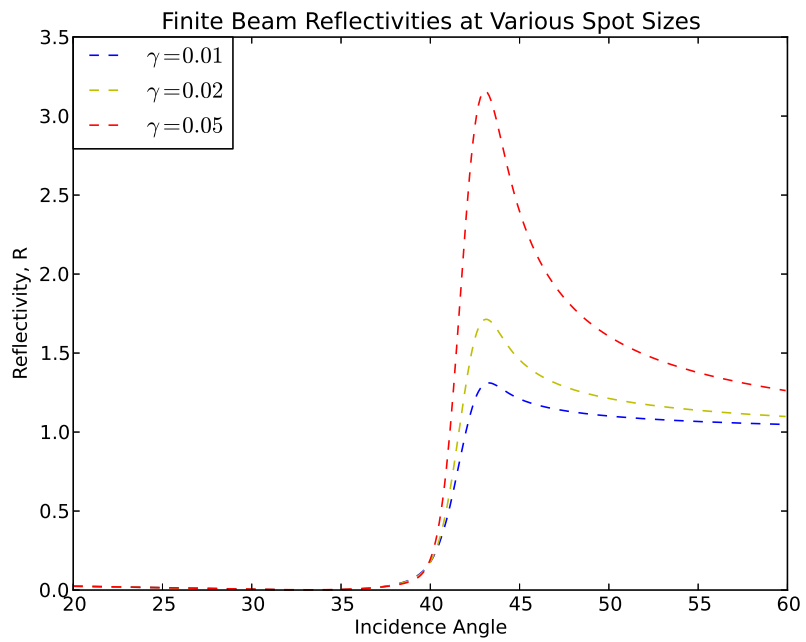


Figure 3-11: Finite beam reflectivities at various gain factors, γ . Beam width fixed at $w_0 = 10\lambda$. At physically realizable gains, increasing γ increases reflectivity.

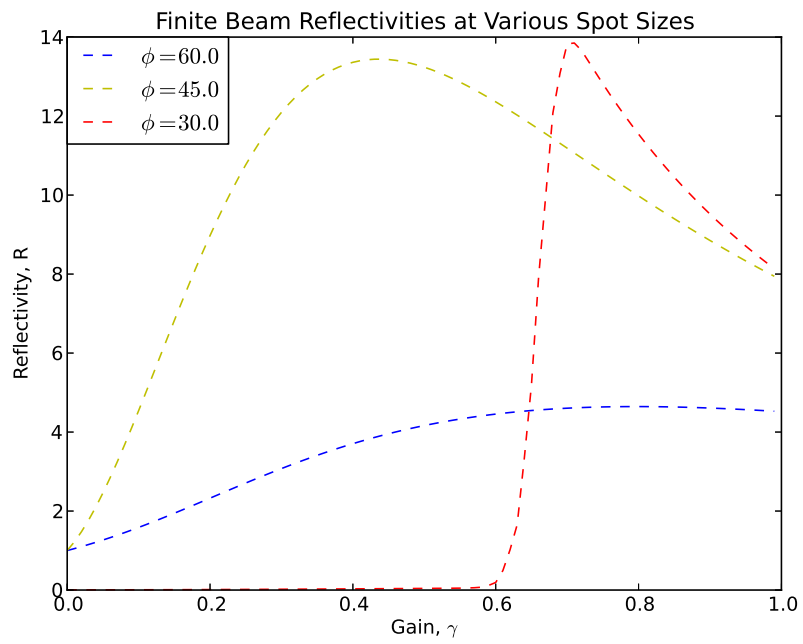


Figure 3-12: Finite Beam Reflectivities at a few incident angles, plotted vs γ . Reflectivity is bounded for a fixed incident angle, and the critical angle moves as γ changes.

Chapter 4

Conclusions

Total internal reflection has a long history of research and practical applications. Amplified total internal reflection from optically gainy media began to be studied, first experimentally and then theoretically in the 1960's. Experimental results have supported the existence of ATIR from the beginning, but difficulties in experimentally studying a single interface led to disagreement among theorists.

By analyzing in detail the case of plane waves incident on the boundary surface of both gain and loss media, this thesis claims to settle the debate about the existence of single surface amplified total internal reflection. Applying well-known boundary conditions to solutions of Maxwell's equations at the gainy or lossy surface leads to reflection coefficients for the cases of s and p polarization respectively given by

$$\tilde{r}_s = \frac{k_{iz} - \tilde{k}_{tz}}{k_{iz} + \tilde{k}_{tz}} \quad (4.1)$$

$$\tilde{r}_p = \frac{\epsilon_1 \tilde{k}_{tz} - \tilde{\epsilon}_2 k_{iz}}{\epsilon_1 \tilde{k}_{tz} + \tilde{\epsilon}_2 k_{iz}} \quad (4.2)$$

These results are agreed upon throughout the literature. The disagreement arises in choosing the value for \tilde{k}_{tz} . For every set of system configurations, \tilde{k}_{tz} can take on one of two mathematically correct conjugate values. Choosing the correct quadrant in the complex plane for \tilde{k}_{tz} dictates whether reflectivity is amplified or attenuated.

By satisfying the following set of three conditions, the thesis shows that amplified total internal reflection occurs from gainy medium beyond the critical angle only.

1. A wave's amplitude must grow in the propagation direction in a gainy medium and decay in a lossy medium.
2. A non-evanescent wave behind the boundary, must propagate away from the boundary in either lossy or gainy media.
3. The solution corresponding to an evanescent wave behind the boundary, must decay away from the boundary in either lossy or gainy media.

Because reflectivity is enhanced at incidence angles beyond the critical angle, but not below it, a discontinuous jump in reflectivity occurs at the critical angle. This jump is surprising because so few quantities in nature exhibit discontinuities. It is significant, however, that the discontinuity is eliminated in the case of a finite width beam.

To analyze the case of a finite width beam, Fourier analysis is used to decompose an incident beam of Gaussian profile into a series of plane waves traveling in slightly different directions. As the incidence angle of the finite beam's central ray is tuned through the critical angle, each the beam's constituent plane waves transitions through its own discontinuity one at a time such that the composite reflectivity is continuous and smooth. The reflectivity of a narrowly focused beam increases slowly and steadily as incidence angle is increased, while that of a wide beam (a better approximation of a plane wave) follows closely to the curve for a plane wave, with a sharp but still smooth rise in reflectivity near the critical angle.

Having resolved the question as to which quadrant is correct for \tilde{k}_{tz} , all reflectivities can be calculated by direct evaluation of the reflection coefficients above. These results are in good agreement with those of FDTD simulations run by other authors, but do not require the same computational power.

References

- [1] R.W. Boyd. *Nonlinear Optics*. Electronics & Electrical. Acad. Press, 2003.
- [2] P. R. Callary and C. K. Carniglia. Internal reflection from an amplifying layer. *Optical Society of America*, 66(8), 1976.
- [3] James W. Cooley and John W. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19:297–301, 1965.
- [4] J. Fan, A. Dogariu, and L. J. Wang. Amplified total internal reflection. *Optics Express*, 11(4):299–308, 2003.
- [5] E. Hecht. *Optics*. Pearson education. Addison-Wesley, 2002.
- [6] John D. Jackson. *Classical Electrodynamics Third Edition*. Wiley, third edition, August 1998.
- [7] Charles J. Koester. Laser action by enhanced total internal reflection. *IEEE Journal of quantum electronics*, 2(9), 1966.
- [8] B.Ya. Kogan, V.M. Volkov, and S.A. Lebedev. Superluminescence and generation of stimulated radiation under internal reflection conditions. *JETP Letters*, 16(3):144–147, 1972.
- [9] Valery N. Konopsky and Elena V. Alieva. Observation of fine interference structures at total internal reflection of focused light beams. *Physical Review*, A86, 2012.

- [10] S.A. Lebedev, V.M. Volkov, and B.Ya. KoganS.A. Lebedev. Value of the gain for light internally reflected from a medium with inverted population. *Optical Spectroscopy*, 35(565), 1973.
- [11] Tobias S. Mansuripur and Masud Mansuripur. The nonexistence of single-surface amplified total internal reflection. Personal Communication, September 2011.
- [12] G. A. Plotz, H. J. Simon, and J. M. Tucciarone. Enhanced total reflection with surface plasmons. *Optical Society of America*, 69(3):419–422, 1979.
- [13] G.N. Romanov and S.S. Shakhidzhanov. Amplification of electromagnetic field in total internal reflection from a region of inverted population. *JETP Letters*, 16(5):298–301, 1972.
- [14] Keely J. Willis, John B. Schneider, and Susan C. Hagness. Amplified total internal reflection: theory, analysis, and demonstration of existence via fdtd. *Optics Express*, 16(3), 2008.

Appendix A

Source Code

The library below is written in python and used to make all figures in this document as well as to perform the discrete Fourier transforms and numerical integrals used in analyzing Gaussian beams. It is released for free use for any purpose. The code relies critically on the numpy, scipy, and matplotlib libraries, and couldn't have been realized without them.

```
1 import numpy as np
2 import numpy.fft as fft
3 from scipy.integrate import quad
4
5 class boundary:
6     def __init__(self, m1, m2real, m2gain, set="n"):
7         if set=="n":
8             self.n1 = m1
9             self.n2 = m2real
10            self.gamma = m2gain
11            self.e1 = m1 ** 2
12            self.e2 = complex(m2real ** 2 - m2gain ** 2, -2 * m2real * m2gain
13                               )
14            elif set=="e":
15                self.e1 = m1
16                self.e2 = complex(m2real, m2gain)
17                #maybe calculate the other n's if necessary
18                self.n1 = np.sqrt(m1)
19            else:
20                raise ValueError("Invalid set type. Must specify one of the
21                                   following:\n n: refractive indicies\n e: permittivities")
22            self.theta_c = np.arcsin(np.sqrt(self.e2.real / self.e1))
23
24    def r(self, theta_i, polarization = "p"):
25        p_iz = self.n1 * np.cos(theta_i)
```

```

24     # Calculate p_tz
25     a = self.e2.real - self.e1 * (np.sin(theta_i))**2
26     b = self.e2.imag
27
28     real_part = np.sqrt(np.sqrt(a**2 + b**2) + a) / np.sqrt(2)
29     imag_part = np.sqrt(np.sqrt(a**2 + b**2) - a) / np.sqrt(2)
30     imag_part = np.where(b < 0, -imag_part, imag_part) # Basically a
        sign function, but treats b=0
31
32     p_tz = real_part + imag_part * 1j
33     p_tz = np.where(self.e2.imag < 0 and theta_i > self.theta_c, -p_tz,
        p_tz)
34
35     if polarization == "p":
36         return (self.e1 * p_tz - self.e2 * p_iz) / (self.e1 * p_tz + self.
        e2 * p_iz)
37     elif polarization == "s":
38         return (p_tz - p_iz) / (p_tz + p_iz)
39     else:
40         raise ValueError("Polarization must be set to either 'p' or 's'")
41
42     return (real_num / denom) + (imag_num / denom) * 1j
43
44     def R(self, theta_i, polarization="p"):
45         return abs(self.r(theta_i, polarization)) ** 2
46
47
48
49     class gaussian_beam:
50     def __init__(self, w0, E0 = 1, wavelength = 1, polarization="p"):
51         # Distance units for all parameters are in wavelengths
52         self.E0 = E0
53         self.w0 = w0
54         self.wavelength = wavelength
55         self.polarization = polarization
56         self.rayleighRange = np.pi * w0 ** 2 / wavelength
57         self.k = 2 * np.pi / wavelength
58         self.divergence = wavelength / (np.pi * w0)
59
60     def w(self, z, approx = False):
61         return self.w0 * np.sqrt(1 + (z / self.rayleighRange) ** 2)
62
63     def curvatureRadius(self, z):
64         z = np.where(z != 0, z, float("inf")) # To avoid division by zero
        when z=0
65         return z + self.rayleighRange ** 2 / z
66
67     def gouy(self, z):
68         return np.arctan(z / self.rayleighRange)
69
70     def complexParameter(self, z):
71         return z + self.rayleighRange * 1j # z + iz_R
72

```

```

73 def E(self, x, y, z, approx = False):
74     rsq = x ** 2 + y ** 2
75     if approx:
76         exponent = -rsq/(self.w0 ** 2) + 1j * self.k * z + 1j * z / self.
            rayleighRange
77         coefficient = self.E0
78     else:
79         exponent = -rsq/(self.w(z) ** 2) + 1j * self.k * (z - rsq/(2 *
            self.curvatureRadius(z))) + 1j * self.gouy(z)
80         coefficient = self.E0 * self.w0/self.w(z)
81
82     return coefficient * np.exp(exponent)
83
84 class gaussian_reflection:
85     def __init__(self, i_beam, boundary, phi = None):
86         self.i_beam = i_beam
87         self.boundary = boundary
88         self.phi = phi if phi is not None else boundary.theta_c
89
90     def transform_analytical(self, theta_i):
91         coefficient = self.i_beam.E0 * self.i_beam.w0 * np.sqrt(np.pi) / np.
            cos(self.phi)
92         exponent = -self.i_beam.k ** 2 * self.i_beam.w0 ** 2 * (np.sin(self.
            phi) - np.sin(theta_i)) ** 2 / (4 * np.cos(self.phi) ** 2)
93         return coefficient * np.exp(exponent)
94
95     def E_r_analytical(self, x, z):
96         if np.atleast_1d(x).size != np.atleast_1d(z).size:
97             if np.atleast_1d(x).size == 1:
98                 x = x * np.ones(z.size)
99             elif np.atleast_1d(z).size == 1:
100                 z = z * np.ones(x.size)
101             else:
102                 raise ValueError('x and z must have same first dimension')
103
104         E = np.empty(0) #should be len(x) maybe
105         for n in range(0, np.atleast_1d(x).size):
106             real_part = quad(lambda theta: \
107                 (self.transform_analytical(theta) * self.boundary.r(theta, self.
108                     i_beam.polarization) \
109                 * np.exp(1j*self.i_beam.k*(x[n]*np.sin(theta)+z[n]*np.cos(theta)))
110                 ).real \
111                 ,0, np.pi/2)
112             imag_part = quad(lambda theta: \
113                 (self.transform_analytical(theta) * self.boundary.r(theta, self.
114                     i_beam.polarization) \
115                 * np.exp(1j*self.i_beam.k*(x[n]*np.sin(theta)+z[n]*np.cos(theta)))
116                 ).imag \
117                 ,0, np.pi/2)
118
119             E = np.append(E, complex(real_part[0], imag_part[0]))
120         return E

```

```

118
119 def transform_fft(self, N = 2 ** 8):
120     self.N = N
121     self.xmax = N * self.i_beam.wavelength/4
122     self.x = np.linspace(-self.xmax, self.xmax, N)
123
124     # Transform the i_beam and store member variables
125     xpb = self.x * np.cos(self.phi)
126     zpb = self.x * np.sin(self.phi)
127     self.E_i = self.i_beam.E(xpb, 0, zpb)
128     self.F_i = fft.fft(self.E_i)
129     self.kx = 2 * np.pi * fft.fftfreq(N, .5)
130     self.kz = np.sqrt(self.i_beam.k ** 2 - self.kx ** 2)
131     self.theta_i = np.arcsin((self.kx * self.i_beam.wavelength) / (2 *
        np.pi))
132
133     # Multiply by reflection coefficients
134     self.r = self.boundary.r(self.theta_i, self.i_beam.polarization)
135     self.F_r = self.r * self.F_i
136
137     # Surface inverse transform
138     self.E_r_surface = fft.ifft(self.F_r)
139
140 def E_r_fft(self, x, z, unprimed_coordinates = True):
141     self.kxpp = -self.kx * np.cos(self.phi) - self.kz * np.sin(self.phi)
142     self.kzpp = self.kx * np.sin(self.phi) - self.kz * np.cos(self.phi)
143
144     if np.atleast_1d(x).size != np.atleast_1d(z).size and np.atleast_1d(
        x).size != 1 and np.atleast_1d(z).size != 1:
145         raise ValueError('x and z must have same first dimension')
146
147     (kx, kz) = (self.kx, self.kz) if unprimed_coordinates else (self.
        kxpp, self.kzpp)
148
149     E = np.empty(0) #should be len(x) maybe
150     for thisx in np.nditer(x):
151         exponent = 1j * (kx * thisx + kz * z)
152         E = np.append(E, np.sum(self.F_r * np.exp(exponent)))
153
154     return np.fft.fftshift(E)/np.sqrt(self.N*x.size)

```