"Superluminal" Pulse Propagation Dynamics

by

George M. Gehring

Submitted in Partial Fulfillment of
the requirements for the Degree
Doctor of Philosophy

Supervised by
Professor Robert W. Boyd

The Institute of Optics
Arts, Sciences and Engineering
Edmund A. Hajim School of Engineering and Applied Sciences

University of Rochester
Rochester, New York

2011

To Barbara

Curriculum Vitae

George Gehring was born in Reading, Pennsylvania in 1981. He graduated second in his class from Exeter Township Senior High School in 1999. He attended Juniata College, receiving a Bachelor of Science degree in Physics and graduating *Summa Cum Laude* with distinction in 2003. During his undergraduate career, he participated in several research projects, including the construction of a laser diode controller unit, manipulation of polystyrene microspheres and human red blood cells with an optical tweezers setup, and statistical analysis of a proposed molybdenum-based neutrino detector.

In 2003, he joined the Institute of Optics at the University of Rochester to pursue a graduate degree. Under the direction of his advisor, Prof. Robert W. Boyd, he studied a variety of topics in the areas of slow- and fast-light propagation and quantum interference. These include propagation dynamics and distortion management in fiber-based slow-light systems, evanescent-field biosensing with ring resonator structures, single-photon quantum interference effects, and single-photon tunneling delays through one- and two-dimensional barrier structures.

Publications related to the thesis

1. "Observation of Backwards Pulse Propagation Through a Medium with a Negative Group Velocity," G. M. Gehring, A. Schweinsberg, C. Barsi, N. Kostinski and R. W. Boyd, *Science.* **312**, 895 (2006).

2. "Reducing pulse distortion in fast-light pulse propagation through an erbium-doped fiber amplifier," H. Shin, A. Schweinsberg, G. Gehring, K. Schwertz, H. Chang, R. W. Boyd, Q-H. Park and D. J. Gauthier, *Opt. Lett.* **32**, 906–908 (2007).

3. "Fiber-Based Slow-Light Technologies," G. M. Gehring, R. W. Boyd, A. L. Gaeta, D. J. Gauthier, and A. E. Willner *J. Light. Tech.*, **26**, 3752–3762 (2008).

4. "Tunable optical time delay of quantum signals using a prism pair," G. M. Gehring, H. Shin, R. W. Boyd, C-M. Kim, and B. S. Ham, *Opt. Express* **18**, 19156–19162 (2010).

Other publications

1. "Coupled-resonator-induced transparency in a fiber system," D. D. Smith, N. N. Lepeshkin, A. Schweinsberg, G. Gehring, R. W. Boyd, Q-H. Park, H. Chang and D.J. Jackson, *Opt. Comm.*, **264**, 163–168 (2006).

Conference papers related to the thesis

1. "Single-Photon Tunneling Delay in a Nematic Liquid-crystal Frustrated-Total-Internal-Reflection Structure," G. M. Gehring, A. C. Liapis, S. G. Lukishova, and R. W. Boyd, *Conference on Lasers and Electro-Optics (CLEO)*, QThN6 (2011).

2. "Backwards Pulse Propagation in Erbium Doped Fiber," G. M. Gehring, A. Schweinsberg, H. Shin, and R. W. Boyd, *Conference on Coherence and Quantum Optics (CQO)*, CSuA (2007).

3. "Pulse Propagation with a Negative Group Velocity in Erbium Doped Fiber," G. M. Gehring, A. Schweinsberg, H. Shin, and R. W. Boyd, *OSA Frontiers in Optics (FiO)*, FThG (2006).

4. "Backwards Pulse Propagation with a Negative Group Velocity in Erbium Doped Fiber," G. M. Gehring, A. Schweinsberg, C. Barsi, N. Kostinski, and R. W. Boyd, *OSA Slow and Fast Light (SL)*, WA2 (2006).

5. "Observation of Exotic Pulse Propagation in Erbium-Doped Optical Fibers," G. M. Gehring, A. Schweinsberg, C. Barsi, N. Kostinski, and R. W. Boyd, *Conference on Lasers and Electro-Optics (CLEO)*, CThW (2006).

Other conference papers

Acknowledgments

A dissertation is a quest for knowledge fueled by the personal desire to discover and understand more about the nature of the world. It is an intimately personal experience, and a test of one's resolve and limitations. However, it would be a pursuit doomed to failure without a great deal of assistance; an individual achievement made possible only through the contributions of many. I gratefully acknowledge the help of those individuals who have contributed to my success in this venture.

I appreciate the guidance and support of my thesis advisor, Professor Robert W. Boyd. The enthusiasm for research that he exhibits has been instrumental in my development as a scientist and encouraged me to press onward even when the path itself was not clear. His natural curiosity and inquisitiveness is a constant reminder that around every corner there is something new to discover.

I would like to thank the faculty members of the Institute of Optics, from whom I learned a great many things. In particular I wish to acknowledge Prof. Miguel Alonso, Prof. Carlos R. Stroud Jr., Prof. Andrew J. Berger, and Prof. James R. Fienup for their mentoring and for interesting discussions on a variety of topics. I am thankful for all of the staff members who have contributed to the success of my graduate career, especially Maria Schnitzler, Nolene Votens, Per Adamson, Joan Christian, Lissa Cotter, Gina Kern, and Marie Banach.

I would also like to express gratitude to Prof. Herbert G. Winful at the University of Michigan for his insights and discussions regarding the nature of tunneling. I am indebted to Dr. Svetlana Lukishova for sharing her expertise with liquid crystals and for numerous helpful discussions. I would also like to thank Prof. John C. Howell and Prof. Andrew N. Jordan for agreeing to serve on my thesis committee. I thank Prof. Byoung Seung Ham and Prof. Q-Han Park for collaborations and advice that led to published articles. And I would like to thank Prof. Jamie White at Juniata College for encouraging me to continue my education in Physics at the graduate level.

I owe a great deal of thanks to the former and current students of Prof. Boyd's

research group for fruitful collaborations and discussions, including Dr. Sean Bentley, Dr. Ryan Bennik, Dr. Vincent Wong, Dr. Matt Bigelow, Dr. Ksenia Dolgaleva, Dr. Giovanni Piredda, Aaron Schweinsberg, Colin O'Sullivan, Petros Zerom, Dr. Anand K. Jha, Dr. Zhimin Shi, Dr. Heedeuk Shin, Joseph E. Vornehm, Jr., Andreas C. Liapis, Mehul Malik, Bosheng Gao, and Brandon Rodenburg. I am also grateful for interactions with the postdocs that have worked with our group in the past, especially Dr. Nick Lepeshkin and Dr. Kam Wai Clifford Chan.

I acknowledge the help and support of all of my fellow graduate students at the University of Rochester. I especially want to thank Dr. Rolf Saager, Dr. Nick Usechak, Dr. Nicole Moore, Matt Moore, Dr. Blair Unger, Dr. John Lesoine and Luke Bissel.

Finally, I would like to thank my friends and family for their unconditional support and encouragement.

Abstract

With the recent rediscovery of slow- and fast-light effects, there has been much interest in pulse propagation phenomena that lead to apparently superluminal pulse transmission. This raises immediate questions about causality and special relativity violations and elicits curiosity about the physical mechanisms responsible for these effects. This thesis will address these issues theoretically and experimentally by investigating two such effects.

The first effect we consider is pulse propagation under conditions of negative group velocity in erbium-doped fiber. Predictions suggest that upon propagation through a region of negative group velocity the transmitted pulse exits the region before the peak of the incident pulse enters. In addition, theory predicts the presence of a peak within the region that propagates backwards, linking the incident and transmitted peaks. We determine the accuracy of these predictions experimentally and uncover the physical mechanism that leads to the effect.

We also examine the saturation of photonic tunneling delays with barrier opacity, which is known as the Hartman effect. In particular we address the case of frustrated-total-internal-reflection, which is a two-dimensional tunneling effect. Theoretical treatments of this phenomenon have been limited to special cases in the past. We demonstrate a new method of decomposition that gives a continuous expression for the predicted delay, and present experimental measurements of that delay in a barrier structure constructed from glass prisms and a liquid crystal cell.

Contents

List of Figures

Foreword

This thesis is the culmination of several research projects I have performed during my doctoral studies at the University of Rochester. All of these projects were collaborative efforts that included my advisor, fellow group members, and occasionally colleagues from other research groups or Universities.

Chapter 1 discusses the paradox of superluminality and provides the motivation for our studies.

Chapter 2 introduces slow- and fast-light propagation. It presents a review and analysis of previous work in the field, and contains fundamental concepts which are necessary to understand the "backwards propagation" effects described in the next chapter.

Chapter 3 is based on a research paper entitled "Observation of Backwards Pulse Propagation Through a Medium with a Negative Group Velocity" published in *Science*, **312**, 895 (2006). This paper was the result of a joint effort by myself, A. Schweinsberg, Prof. R. W. Boyd, and undergraduate research students C. Barsi and N. Kostinski. I conceived the idea for the project, designed the experimental setup, was heavily involved in the data collection and analysis, and wrote the paper for publication. C. Barsi and N. Kostinski assisted with preliminary data collection. A. Schweinsberg was heavily involved in data collection and analysis, and both A. Schweinsberg and R. W. Boyd were involved in the discussion and revision of the manuscript. I am the first author of the published paper.

Chapter 4 contains the derivation of one-dimensional tunneling delay in both

the "phase time" and "dwell time" interpretations. This derivation is heavily based on several previous works by other authors, though I have collected and arranged these works so as to present a more complete derivation with a unified notation. This chapter contains the necessary foundation upon which we will construct our derivation for the two-dimensional case.

Chapter 5 contains the complete derivation of the two-dimensional tunneling delay and Goos-Hänchen shift for a symmetric system comprised of two different materials of indices n_1 and n_2. It is unpublished work performed by myself with direction and guidance from R. W. Boyd. It has, however, benefitted greatly from conceptual discussions with H. G. Winful. We intend to prepare this work for publication in the near future.

Chapter 6 describes the double prism system we utilized to measure tunneling delay in transmission and reflection, provides the results of numerical simulations and experimental measurements, and analyzes the experimental results. It is also unpublished work that was primarily performed by myself with direction and guidance from R. W. Boyd. However, the construction and operation of the experimental setup was influenced and expedited through the assistance of several graduate students, particularly A. K. Jha, H. Shin, C. O'Sullivan, P. Zerom, and A. C. Liapis. We intend to prepare this work for publication in the near future.

Chapter 1

Introduction

The topic of superluminal transit has been hotly debated since Einstein's original suggestion of a cosmic speed limit. It is a common misconception that the special theory of relativity prohibits velocities in excess of the speed of light. In fact, the special theory of relativity only prohibits the transfer of information faster than c, though a formal definition of what constitutes information has never been determined [1]. However, most reasonable definitions include matter, energy, and any signal that can be used deterministically, such as the presence or absence of a light pulse in a time or bit slot.

The common consensus among modern scientists is that superluminal information transit is unphysical, but there are many physical phenomena that allow the peak of a pulse to propagate faster than the speed of light. With the resurgence of interest in slow- and fast-light propagation in recent years, there is a large body of experimental data verifying that under certain conditions pulses can indeed appear to travel in ways that seem acausal to an external viewer. These experiments are often poorly reported in the mainstream media, giving the impression that superluminal transmission of information is possible.

However, superluminal information transfer has never been demonstrated. The "superluminal" velocities reported are always inferred based on the peak of a spec-

trally and temporally broad waveform, and under those conditions the peak cannot be said to carry any information. Despite this fact, it is commonplace for researchers to report experimental results describing the advancement of a pulse peak in terms of "superluminal" velocities.

In most cases, academic researchers are careful to note that their measurements of pulse velocity, group velocity, or transit time are not subject to the same causality constraints that a true signal velocity would be. Other authors have addressed the issue head-on by performing experiments to demonstrate that such velocities do not lead to superluminal signal transit [2]. However, there have been a number of recent publications where results have been misinterpreted by the authors, leading to erroneous claims that they have observed superluminal pulse propagation effects which violate special relativity [3–7].

It is our goal to investigate these apparently superluminal effects by examining their underlying propagation dynamics. By studying the details of the pulse propagation within the material, we feel it is possible to discern and understand the physical mechanism that leads to the effect. In this fashion, the cause of the apparent superluminality can be uncovered and the ambiguity leading to misinterpretation can be removed. In short, we seek to identify the *causal* mechanisms that lead to the appearance of *acausal* pulse behavior. In the remainder of this thesis, we will perform this investigation for two of the most common physical phenomena that lead to "superluminal" pulse propagation.

The next two chapters will discuss slow- and fast-light propagation, with a focus on negative group velocities and the "backwards propagation" effect that is predicted under those conditions. When propagating through a negative-group-velocity material, theoretical models suggest that the peak of the transmitted pulse actually exits the negative-group-velocity material before the peak of the incident pulse enters it. This also suggests that the optical field within the material exhibits a peak which travels *backwards* from the exit to the entrance, linking the incident and transmit-

ted pulses. These strange predictions lead to questions about causality and energy conservation which can be addressed experimentally.

Chapters 4-6 will address time delays in quantum-mechanical barrier tunneling and the Hartman effect. When a particle tunnels through a potential barrier, the peak of the transmitted particle's wavefunction may appear earlier than a causally-propagated copy of the incident particle's wavefunction would. This leads to the interpretation that the time delay in tunneling is superluminal in nature and demonstrates a violation of special relativity. There are electromagnetic analogs of this effect in which photons exhibit similar behavior. We will demonstrate theoretically that it is incorrect to interpret these delays as propagation times, and instead propose a cavity interpretation of tunneling that eliminates the paradox. We will also provide theoretical predictions and experimental measurements of the expected propagation delay in a two-dimensional electromagnetic tunneling setup.

Chapter 2

Slow and fast light

The study of light, particularly its velocity and propagation behavior, goes back hundreds of years. However, the terms "slow and fast light" are recent concoctions that refer to a very specific pulse propagation phenomenon. In this chapter, we will briefly overview the theory of slow- and fast-light propagation and discuss some of the important points that will be relevant to our experimental analysis of "backwards propagation" in chapter 3.

2.1 Phase velocity and group velocity

The simplest form of an electromagnetic wave is a "plane wave" in which the phase fronts are planes that propagate in a direction normal to the plane. A plane wave can be expressed mathematically in space and time $(\mathbf{r}, t)$ as

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{e}E_0 e^{i\mathbf{k}\cdot\mathbf{r} - i\omega t} \tag{2.1}$$

where $\mathbf{e}$ is a vector describing the direction of the electric field (or equivalently, the polarization), E_0 is the amplitude of the wave, $\mathbf{k}$ is the wavevector or propagation vector, and ω is the frequency of the light. In free space, $\mathbf{e} \cdot \mathbf{k} = 0$, indicating that

the light is always polarized perpendicularly to the direction of propagation. Thus we can arbitrarily define the direction of propagation to be the z-direction, and simplify equation (2.1):

$$\mathbf{E}(z,t) = \mathbf{e}E_0 e^{ikz - i\omega t}. \tag{2.2}$$

In this form, k is the wave number or magnitude of the wavevector ($\mathbf{k} = k\hat{z}$), and $\mathbf{e}$ is constrained to be in the x-y plane. For the moment, we will assume an ideal lossless material and thus consider only the real part of $\mathbf{k}$. We will address the implications of loss in the next section.

If we wanted to describe the speed at which a phase front of this wave propagates through space, we need only consider the phase $\phi(z,t) = kz - \omega t$. k and ω are constant, so if we want to determine the motion of a phase front, we can simply differentiate ϕ with respect to t and set the result equal to zero,

$$\begin{aligned} \frac{d\phi(z,t)}{dt} &= k\frac{dz}{dt} - \omega = 0 \\ \frac{dz}{dt} &= \frac{\omega}{k}. \end{aligned} \tag{2.3}$$

In free space, $k = k_0 = \omega/c$, and the phase front moves at the speed of light c as expected. However, in linear media the wave number has an additional factor n, the index of refraction of the material. In the general case, n is also a frequency-dependent quantity, such that $k = n(\omega)\omega/c$. Substituting this into equation (2.3) gives us a general expression for the velocity of the phase front or *phase velocity* $v_p = dz/dt$,

$$v_p = \frac{dz}{dt} = \frac{\omega}{n(\omega)\omega/c} = \frac{c}{n(\omega)}. \tag{2.4}$$

The phase velocity is the appropriate representation of the velocity of a monochromatic wave of light traveling through a medium of refractive index $n(\omega)$.

However, in the laboratory we are frequently more interested in optical pulses of light and time-of-flight measurements. While an optical pulse is made up of a nar-

rowband collection of monochromatic beams, the velocity of the pulse is not generally the phase velocity, because each of the constituent monochromatic beams experiences a slightly different refractive index $n(\omega)$ and thus propagates at a slightly different speed. We can formally develop this thought by applying a time-varying envelope $A(t)$ to our monochromatic wave in equation (2.2),

$$E(z,t) = A(t)e^{ikz-i\omega_0 t}, \tag{2.5}$$

where we have omitted polarization for the time being. The modulation $A(t)$ describes the slowly-varying pulse envelope, while the exponential describes the high-frequency carrier wave at ω_0. In the frequency domain, $A(t)$ is responsible for new frequency components of the pulse on either side of ω_0. Using standard Fourier theory, we can express $E(z,t)$ in the frequency domain explicitly,

$$E(z,t) = \frac{e^{-i\omega_0 t}}{2\pi}\int_{-\infty}^{\infty} A(\omega)e^{ikz-i\omega t}, \tag{2.6}$$

in which the factor $A(\omega)$ is the Fourier transform of $A(t)$. Written in this form it is clear that the peak of the pulse, or the maximum value of $A(t)$, should occur where all of the constituent frequency components interfere constructively. This will occur where the phase of exponential in the integrand $\phi = kz - \omega t$ does not vary with frequency, or $d\phi/d\omega = 0$. This is also known as a "stationary phase" approximation, a common technique for evaluating oscillatory integrals [8]. If we explicitly carry out this derivative, we can determine the distance z that the peak of the pulse travels in a time t, commonly known as the *group velocity* v_g:

$$\begin{aligned} \frac{d\phi}{d\omega} &= \frac{dk}{d\omega}z - t = 0 \\ v_g &= \frac{z}{t} = \frac{d\omega}{dk}. \end{aligned} \tag{2.7}$$

We can write this in a more explicit form by evaluating the derivative $dk/d\omega$,

$$\begin{aligned}\frac{dk}{d\omega} &= \frac{d}{d\omega}\left(\frac{n(\omega)\omega}{c}\right) \\ &= \frac{n(\omega)+\omega\frac{dn(\omega)}{d\omega}}{c},\end{aligned} \tag{2.8}$$

and substitution into (2.7) gives

$$\begin{aligned}v_g &= \frac{d\omega}{dk} = \frac{1}{\frac{dk}{d\omega}} = \frac{c}{n+\omega\frac{dn}{d\omega}} \\ &= \frac{c}{n_g}\end{aligned} \tag{2.9}$$

where we have defined the *group index* $n_g = n + \omega\frac{dn}{d\omega}$ and suppressed the frequency dependence of n for clarity. The second term of n_g is the feature that leads to slow- and fast-light phenomena. In a dispersionless material, n is constant and $\frac{dn}{d\omega} = 0$, and the phase and group velocities take on the same value. This is the case in vacuum, for example, where n is identically 1. Air is very nearly dispersionless at standard temperature and pressure, leading to similar behavior.

However, in a region where the dispersion $\frac{dn}{d\omega}$ becomes large, the second term of n_g can become significant and play an important role in pulse propagation dynamics. "Slow light" occurs in regions of large normal dispersion, where $\frac{dn}{d\omega} > 0$ and $\omega\frac{dn}{d\omega} > 1$. In many cases, this dispersive contribution dominates, leading to group indices in the hundreds, thousands, or higher. Slow light has been observed in a large variety of systems, including atomic vapors [9–11], Bose-Einstein condensates [12, 13], room-temperature solids [14, 15], and microscale photonic structures [16–20].

Regions of anomalous dispersion cause $\frac{dn}{d\omega}$ to become negative, which can lead to "fast light" effects. If the medium is only weakly dispersive, v_g may simply exceed c, and the pulse will appear to propagate faster than the speed of light. In regions of strong anomalous dispersion, the second term may actually become larger in magnitude than n and force the group velocity to become negative. Negative group velocity

propagation has been predicted by Brillouin and others [21–25], and observed experimentally in a variety of materials [15, 26–31]. In all of these cases, the authors stress that the resulting pulse propagation dynamics are strictly causal, and cannot produce superluminal transmission of information.

2.2 Creating slow and fast light

One common goal within the slow-light research community has been to achieve controllable and tunable slow light. An ideal slow-light material that provided tunability over a high dynamic range for a broad signal bandwidth would open the door towards slow-light optical buffers that may improve the operation of telecommunications equipment, especially if that slow-light was achievable in one of the standard telecommunications bands. Engineering an ideal slow light material is a non-trivial matter, however. The dispersion of the material must be carefully crafted from one of several mechanisms, each of which have their advantages and disadvantages.

In general, the mechanisms that cause slow-light effects can be categorized as either "atomic slow light" or "structural slow light." These distinctions describe the source of the dispersion that causes the slow-light effect. Atomic slow light materials achieve dispersion by exploiting atomic or molecular resonances. Structural slow light occurs when a region of high dispersion is caused by geometric or structural factors, as is the case with waveguide dispersion or photonic band-gap effects.

Atomic slow light can be understood by examining a simple model of absorption and gain. Let us consider the single Lorentzian gain feature shown in Figure 2.1a, which has the form

$$g(\Delta\omega) = \frac{g_0}{1 + \Delta\omega^2/\gamma^2}, \tag{2.10}$$

where $\Delta\omega = \omega - \omega_0$ is the detuning from the resonance frequency ω_0, γ is the linewidth of the resonance, and g_0 is the value of the intensity gain coefficient at line center. This form could model a collection of atoms in an excited state that transitions to

the ground state through the emission of a photon at frequency ω_0.

The frequency-dependent optical intensity absorption coefficient $\alpha(\omega)$, which is the opposite of $g(\omega)$ in our model, is linked to the refractive index of a material through a set of Hilbert transforms known as the Kramers-Kronig (KK) relations [32, 33],

$$n(\omega) = 1 + \frac{c}{\pi} \int_0^\infty \frac{\alpha(\Omega)}{\Omega^2 - \omega^2}\, d\Omega, \tag{2.11a}$$

$$\alpha(\omega) = -\frac{4\omega^2}{\pi c} \int_0^\infty \frac{n(\Omega) - 1}{\Omega^2 - \omega^2}\, d\Omega. \tag{2.11b}$$

Note that the dashed integrals indicate a Cauchy principal value. These relations are an outcome of causality [34] and are fundamental to atomic slow light effects. They predict that for our Lorentzian gain feature, we will have a refractive index profile shown in Figure 2.1b of the form

$$n(\Delta\omega) = n_0 + \left(\frac{g_0 c}{2\omega_0}\right) \frac{\Delta\omega/\gamma}{1 + \Delta\omega^2/\gamma^2}, \tag{2.12}$$

where n_0 is the real frequency-independent "background" index of refraction. From $n(\Delta\omega)$, we can calculate the group index $n_g(\Delta\omega)$ shown in Figure 2.1c,

$$n_g(\Delta\omega) = n(\Delta\omega) + \left(\frac{g_0 c}{2\gamma}\right) \frac{1 - \Delta\omega^2/\gamma^2}{(1 + \Delta\omega^2/\gamma^2)^2}. \tag{2.13}$$

At line center, we have a region of strong normal dispersion which leads to a large slow light effect. The group index n_g is much greater than n_0 if $g_0 L/\gamma$ is large. The expected "group delay" for a narrowband pulse centered at ω_0 propagating through this medium is $\tau_g = n_g L/c = n_0 L/c + g_0 L/2\gamma$. In the wings of the gain resonance, we have weak anomalous dispersion and observe a region of fast light, where the group index becomes negative.

Since the Kramers-Kronig relations are a pair of Hilbert transforms, they are linear functions. Thus, if we invert the sign of g to create an absorption resonance, the Kramers-Kronig relations preserve the sign change and predict the opposite effect. In

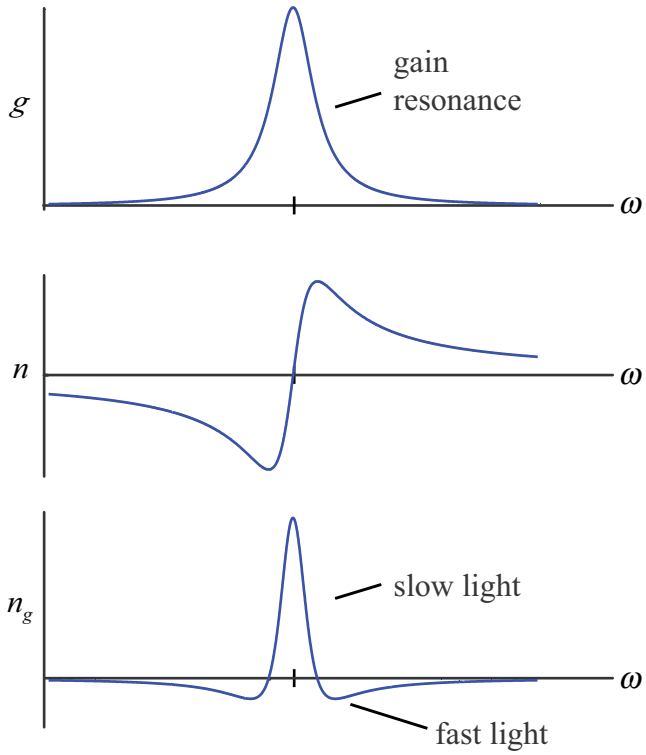

Figure 2.1: Lorentzian gain peak and the dispersion profiles of the refractive index n and group index n_g associated with this gain spectrum by the Kramers-Kronig relations. It is clear from the figure that a very large and positive group index occurs in the region of strong linear dispersion in n. This region of slow-light propagation is flanked by fast-light regions due to strong normal dispersion in the wings of the refractive index dispersion profile.

that case, we see a strong fast-light effect on line center and a weak slow-light effect in the wings. In addition, linearity forces them to be insensitive to a constant background value, so they give the same prediction for an absorption resonance described by $-g(\omega)$ and a dip in a background described by $g_0 - g(\omega)$. This will be the particular situation of study in our experiment, where we investigate fast light caused by a narrow absorption resonance occurring within a broadband gain feature.

Atomic slow light accurately describes the pulse propagation effects observed from single gain- or absorption-line structures [35–39] and complicated multiple-gain-line or multiple-absorption-line structures [40, 41]. It also includes nonlinear processes

[42] such as electromagnetically-induced transparency (EIT) [9, 12, 43, 44], four-wave mixing (FWM) [45], and coherent population oscillations (CPO) [14, 15, 31, 46–49].

In structural slow light, it is not the absorption or refractive index that produces the desired dispersion. Instead, these systems support propagating modes whose group velocity is governed by the geometry or structure of the system. Note that in general, these effects are still bound by a Kramers-Kronig relation, as they can be described by a complex transfer function $\tilde{H}(\omega) = A(\omega)\exp[i\phi(\omega)]$ that is based on a causal temporal response function. However, in this case, the Kramers-Kronig relation couples the amplitude and phase of the frequency response function rather than absorption and refractive index, and may be more complicated to express analytically.

As a simple example, consider a waveguide that supports one or more mode in some frequency range. The modes have a frequency-dependent "effective wavevector" β that depends on the confinement and size parameters of the waveguide. These are frequently characterized by a dispersion curve $\beta(\omega)$, from which group velocity can be easily calculated. Adjustment of the waveguide size parameters, index contrast, or geometry all have an effect on β and can be used to control the dispersion characteristics and subsequently the group velocity of the waveguide. Structural slow light also encompasses band-gap effects, or regions where propagating modes aren't supported. Near the edge of these regions β becomes highly dispersive, leading to significant deviations in the group velocity.

"Designer" waveguides constructed in photonic crystal structures have demonstrated slow-light effects [17, 50, 51], as have fiber Bragg grating structures [52–54]. Even in weakly dispersive waveguides, one can create slow light effects by introducing an external source of dispersion. This is frequently accomplished by introducing a resonator of some sort that couples to the waveguide, creating the structural equivalent of an atomic resonance that modifies the dispersion properties of the system [55–59]. One can even construct the waveguide entirely out of these resonator structures [19, 20, 60–62]. Alternatively, one can combine structural and atomic tech-

niques by introducing a true atomic resonance to the waveguide geometry [39,63,64].

2.3 Practical considerations

In practice, there are a number of complications that must be dealt with in any slow-light system. The first of these is the effect of higher-order dispersion on the pulse. To demonstrate these effects, we expand the real part of the wavevector $k(\omega)$ (or equivalently $\beta(\omega)$) using a Taylor series,

$$\begin{aligned} k(\omega) &= k(\omega_0) + \left.\frac{dk}{d\omega}\right|_{\omega_0}(\omega-\omega_0) + \left.\frac{1}{2!}\frac{d^2k}{d\omega^2}\right|_{\omega_0}(\omega-\omega_0)^2 + \left.\frac{1}{3!}\frac{d^3k}{d\omega^3}\right|_{\omega_0}(\omega-\omega_0)^3 + ... \\ &= k(\omega_0) + \beta_1\Delta\omega + \frac{\beta_2}{2!}\Delta\omega^2 + \frac{\beta_3}{3!}\Delta\omega^3 + ... \end{aligned} \tag{2.14}$$

where

$$\beta_n \equiv \left.\frac{d^n k}{d\omega^n}\right|_{\omega_0} \tag{2.15}$$

is the n^{th}-order dispersion factor. β_1 is simply $n_g(\omega_0)/c$, or $1/v_g$. In the ideal case where all higher-order dispersion factors are zero, $k(\omega) = k(\omega_0) + \beta_1(\omega-\omega_0)$ and the pulse at position $z = L$ and time t can be described very simply:

$$\begin{aligned} E(L,t) &= \int_{-\infty}^{\infty} d\omega E(L,\omega)e^{-i\omega t} \\ &= \int_{-\infty}^{\infty} d\omega E(0,\omega)e^{ik(\omega_0)L+\beta_1(\omega-\omega_0)L-i\omega t} \\ &= e^{ik(\omega_0)L+i\beta_1\omega_0 L}\int_{-\infty}^{\infty} d\omega E(0,\omega)e^{-i\omega(t-\beta_1 L)} \\ &= e^{ik(\omega_0)L+i\beta_1\omega_0 L}E\left(0,t-\beta_1 L\right) \end{aligned} \tag{2.16}$$

Apart from a phase term, the pulse at $z = L$ looks identical to the pulse at $z = 0$ but shifted in time by $\Delta t = \beta_1 L = n_g L/c$. Thus, the pulse preserves its shape upon propagation and travels at the group velocity.

Higher-order dispersion terms disturb the relative phases of different frequency

components, causing a number of distortion effects. The largest of these is group velocity dispersion (GVD), governed by β_2, which causes pulse broadening or compression. GVD is particularly important in telecommunications applications. If the pulse broadens enough it may encroach on a neighboring bit slot, leading to crosstalk and higher bit-error rates. Third, fourth, and higher-order dispersion terms can lead to more complicated distortion effects, including severe pulse reshaping and break-up.

Frequency-dependent gain or loss can also distort the pulse shape by modifying the signal spectrum. Of course, the higher-order dispersion terms are in no small part due to the frequency variation of gain or loss thanks to the Kramers-Kronig relations. However, dispersion describes pulse distortions caused by phase differences amongst the constituent monochromatic waves. Frequency-dependent gain or loss also causes the amplitude of these monochromatic components to change, which can lead to other pulse distortions. In general, one must consider both the phase and amplitude distortion effects to accurately describe the total pulse distortion.

Furthermore, any loss mechanism will cause the pulse to be diminished in amplitude after propagation through the slow light material. It is often impossible to avoid loss in slow-light systems, which puts a practical limit on the maximum achievable delay and system performance [65]. All of these sources of pulse distortion contribute to whether a particular slow light mechanism is suitable for a given application, or in some cases whether there is *any* slow-light process capable of achieving the desired system performance.

The management or minimization of these effects is an active area of research. It has been demonstrated that the use of multiple gain lines can reduce the amount of distortion by "flattening" the gain curve in the region of the pulse's bandwidth [40]. In some systems, an additive background intensity at the probe frequency can also reduce pulse distortion [66, 67]. For single-line slow light effects like stimulated Brillouin scattering, electronic broadening of the gain line can lead to an increased bandwidth available for slow light, reducing distortion for short pulses [38, 68]. Finally, there are

even techniques to reduce less obvious distortion effects like pattern dependence [69].

2.4 Group velocity and causality

Fast light and "backwards" light, or cases where $v_g > c$ or $v_g < 0$ respectively, appear to run afoul of our understanding of causality. They imply that objects are propagating faster than the speed of light, or in the negative-group-velocity case suggest that an object exits a material before it enters. Both of these situations would be causality violations in the special theory of relativity. Upon careful consideration, we find that this is not the case, and that causality is obeyed in all instances of anomalous pulse propagation.

This question has been studied in some detail, dating back to the early 20th century. In their examination of the propagation of discontinuous electric field envelopes, Sommerfeld and Brillouin defined five different velocities of light [70, 71]. They found that the "front velocity," or the velocity of the discontinuity representing pulse turn-on, could never exceed c. Moreover, they found that no part of the waveform could overtake the discontinuous pulse front. These results led to a reformulation of the special theory of relativity to state that rather than an object, "information" could not travel faster than c, despite the fact that a formal definition of information velocity has never been agreed upon [1].

In 1970, Smith expanded the list of velocities to seven, including a "signal velocity" which is almost identical to the front velocity of Brillouin, an energy velocity, and a "centrovelocity" that describes the propagation of the pulse's energy centroid [72]. All of these velocities are distinct and differ from the standard group velocity definition. More recently, the concept of information velocity has returned to the forefront of discussion, and experiments measuring the propagation velocity of pulses with points of non-analyticity have confirmed that these points propagate at c [2, 30, 73].

One way to see that these superluminal propagation effects must be consistent

with causality is by examining their mathematical source. The group velocity is predicted by the complex linear susceptibility $\chi^{(1)}$, which is defined as the Fourier transform of a causal response function [34]. The Kramers-Kronig relations that link the real and imaginary parts of $\chi^{(1)}$ are thus similarly constrained by causality, even when they predict dispersion features that lead to anomalous group velocities. A similar argument can be made for Maxwell's equations, which obey special relativity and govern the propagation of light through these materials. So despite the unusual group velocities predicted, it is impossible for the results of these calculations to violate causality.

What, then, is the meaning of the group velocity? It does not appear to represent the propagation of information, energy, or pulse centroid. Under some circumstances, it may be comparable or equal to some or any of those physical effects, but it isn't in general. In fact, Smith notes that,

> "By its nature, the group velocity is a mathematical entity which may not have any real physical significance associated with it. There is no physical particle, mass, energy, or signal which necessarily travels at the group velocity."

This statement captures the essence of the group velocity quite well. The group velocity describes the time evolution of a group of frequency components, which in many cases also describes the propagation of the peak of an optical pulse. But that peak is merely a mathematical construct. The peak does not have a physical significance beyond being a point of maximum intensity; it is neither object nor signal. And as we've seen, in situations where higher-order dispersion is present that peak can be destroyed or distorted during propagation. As a result, the peak is not bound by relativity or causality, and it should be of no surprise that a quantity describing that peak gives values that are impossible for a physical entity.

That is not to say that the group velocity is a useless quantity. There are numerous applications where the group velocity is relevant to the performance of a system. For

example, in an optical telecommunications system data is encoded in bits represented by the presence or absence of a pulse during a defined bit slot. Ideally, every pulse is centered in its bit slot. However, during propagation a variety of effects may alter the pulse's position within its bit slot, even leading to encroachment on an adjacent bit slot. Slow- and fast-light effects can help re-center the pulse in its bit slot and reduce crosstalk and error rates, even though they have not altered the information velocity.

2.5 Summary

In this chapter we have given an overview of the fundamentals of slow- and fast-light propagation. A brief review of the mechanisms through which these anomalous group velocities can be obtained was presented, and the practical issues that inevitably occur in a practical slow- or fast-light system were discussed. Finally, we discussed the relevance and meaning of the group velocity as well as its relationship to causality and relativity.

In the next chapter, we will discuss our experimental investigation of fast-light or "backwards" pulse propagation in an erbium-doped fiber amplifier.

Chapter 3

Backwards propagation

In this chapter we will discuss the topic of "backwards" propagation or negative group velocities. We will first provide a brief overview of the conceptual interpretation of negative group velocities, and then detail the erbium fiber system in which we will be investigating the effect. A theoretical model that describes this system will be provided, and our experimental investigation of the effect will be detailed.

3.1 Negative group velocity

As we discussed in the previous chapter, under certain conditions the dispersion of the material can become highly anomalous and lead to negative group velocities. This raises a number of questions about the nature of the process and about what is physically occurring during propagation. A number of authors have addressed the theoretical underpinnings of negative group velocity [21–25,74–77]. The conclusions of these works are demonstrated in Figure 3.1, which is a simple simulation of a negative-group-velocity medium. First, the peak of the transmitted pulse is predicted to exit the rear interface of the material (represented by the red box) before the peak of the incident pulse arrives at the first interface. In addition, there appears to be a pulse propagating *backwards* through the material, such that the peak of the backward-

propagating pulse coincides with the peaks of the entering and exiting pulses.

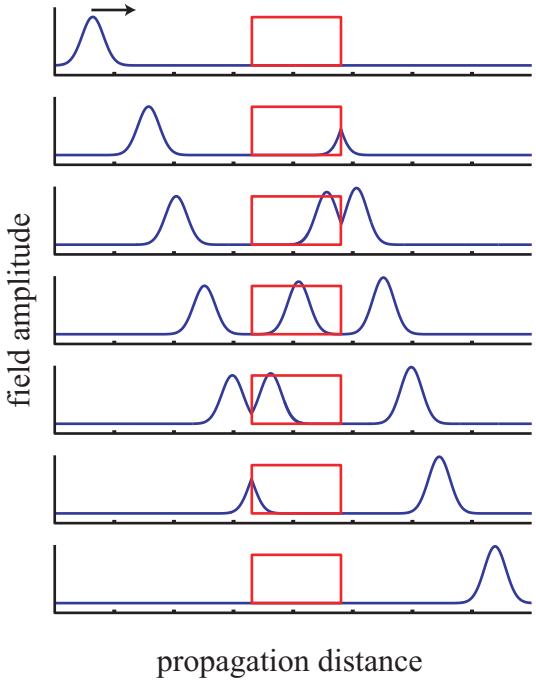

Figure 3.1: Schematic diagram of pulse propagation within a negative group velocity medium. Each panel is a sequential snapshot of the pulse envelope, with the boxes representing the region of negative group velocity. The output pulse peak exits the material between panels two and three, while the input peak does not enter the material until after panel five. In addition, a peak propagates backwards within the material linking output and input pulse peaks.

It should be noted that Figure 3.1 is an idealization that exaggerates the backwards propagation effect for clarity. The simulation assumes a medium with two very strong gain resonances separated in frequency. These gain lines each give a strong slow-light effect at resonance, but combine to form a broad region of negative group velocity between the resonances. The pulse bandwidth is constrained to be within this negative group velocity region, and thus the pulse has infinite temporal extent. If the pulse had any bandwidth component at gain resonance, there would be a considerable amount of GVD and pulse distortion.

Perhaps the most idealized factor is that the simulation does not include any noise

effects, which would be a significant source of error due to the strength of the gain features required to obtain such a significant pulse advancement effect. Any noise component at either gain resonance would be amplified significantly and create intensity fluctuations that would exceed that of the pulse and distort its temporal profile. In a practical backward-propagation system we would expect our pulse advancement to be limited by GVD or noise effects, likely to less than one pulse length.

Pulse advancement in a negative-group-velocity medium has been observed in previous work [15, 26–30, 78]. However, until our contribution, no attempt had been made to investigate the dynamics of the backward-propagating pulse. In fact, it was not entirely clear whether the theoretical predictions of backwards propagation were robust enough to be observed under actual laboratory conditions. In addition, it was unclear whether the energy velocity was positive or negative in this situation. The group velocity and energy velocity are not generally equivalent, and in fact can only be equal in the absence of gain or loss, which is equivalent to abandoning causality and thus un-physical [72].

3.2 CPO in erbium doped fiber

To create our absorption feature, we used a technique known as coherent population oscillations, or CPO. CPO was first demonstrated in a ruby crystal [14, 79], but has also been demonstrated in alexandrite [15], semiconductor quantum wells [46,47], and erbium-doped fiber [31].

The CPO effect occurs when an intensity-modulated probe beam interacts with a homogeneously broadened atomic absorption or gain resonance. The probe causes the atomic ground state population density to oscillate at the frequency of the modulation with a slight lag in the modulation phase. As a result, the pulse sees a time varying absorption or gain, which reshapes the pulse accordingly. For a single pulse, only one oscillation of the population density occurs. The effect can be equivalently understood

as a type of saturable absorption or as a nonlinear effect that creates a narrow spectral hole at the pulse's carrier frequency.

For our experiments, we chose to use erbium-doped optical fiber (EDOF) as our negative-group-velocity material. Erbium has the energy level diagram shown in Figure 3.2, and can act as a saturable absorber or reverse saturable absorber for light at a wavelength of $\lambda = 1550$ nm in the absence or presence of an applied 980-nm pump beam. It is thus a slow-light medium by nature, but becomes a fast-light medium when a pump field is applied.

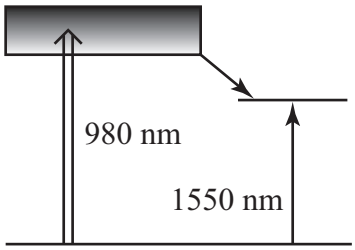

Figure 3.2: Energy level diagram for erbium. A pump applied to the 980-nm transition will invert the 1550-nm transition, changing the system from a saturable absorber to a reverse saturable absorber. This turns erbium from a slow-light medium into a fast-light medium.

In addition to a convenient material system, EDOF gives us the benefits of working in optical fiber. Fiber provides tight confinement and subsequently large optical intensities. It also makes long interaction lengths feasible so that effects can be studied over many meters of distance. Both of these factors enhance nonlinear interactions and lead to stronger slow- and fast-light effects. Finally, the ability to work in fiber at a common telecommunications wavelength allows use of inexpensive and widely-available off-the-shelf equipment.

If we chose to develop our theoretical treatment in the frequency domain, we would follow the method shown in [14]. The probe beam would be represented as a combination of a monochromatic carrier wave and one or more sidebands which

represent the modulation,

$$E(t) = E_0 e^{-i\omega_0 t} + E_1 e^{-i(\omega_0+\delta)t}, \tag{3.1}$$

where ω_0 is the carrier frequency and δ is the intensity modulation frequency. Density matrix analysis would then lead to a modulation of the population inversion and a dip in the atomic absorption profile $\alpha(\delta)$,

$$\alpha(\delta) = \frac{\alpha_0}{1+I_N}\left(1 - \frac{I_N(1+I_N)}{(T_1\delta)^2 + (1+I_N)^2}\right), \tag{3.2}$$

where $I_N = I_0/I_{\text{sat}}$ is the normalized intensity of the central frequency component, I_{sat} is the saturation intensity of the medium, T_1 is the excited state lifetime of the transition, and α_0 is the unsaturated background absorption (or gain) coefficient. The second term in equation (3.2) is the "spectral hole" that causes the slow- or fast-light effects.

However, since the dipole dephasing time T_2 of the medium is much smaller than the excited state lifetime T_1, we may also choose to address the problem in the time domain with a rate equation treatment [31, 80]. This approach provides additional insight into the process, as the effects of the spectral hole are interpreted to be time-dependent saturation of the absorption or gain of the material. The nonlinear saturable-gain picture of the process also nicely complements the purely linear energy exchange description provided in [24]. Since this rate equation formulation is sufficient to describe our experimental results, we will proceed with the time-domain analysis as presented in [31].

We model the erbium atoms as a three-level system shown in Figure 3.2. A pump wave at 980-nm excites the atoms from the ground state into an upper-level state, and a signal wave at 1550-nm experiences the effects of a transition between the ground state and a metastable upper-level state. The decay from the upper-level pumping state to the metastable state is assumed to be rapid compared to T_1, the

metastable state's lifetime. Under those conditions, the rate equation for the ground state population density n can be written,

$$\frac{dn}{dt} = \frac{\rho - n}{T_1} + \left(1 - \frac{n}{\rho}\right)\beta_s I_s - \frac{n}{\rho}\alpha_s I_s - \frac{n}{\rho}\alpha_p I_p, \tag{3.3}$$

where ρ is the Er^{3+} ion density in the fiber, I_p is the pump intensity in units of photons/area/time, I_s is the signal intensity, β_s is the stimulated emission coefficient at the signal wavelength, and α_p and α_s are the absorption coefficients for pump and signal waves, respectively. In this expression, the first term describes the natural decay of the metastable state and the second term describes stimulated emission caused by the signal wave. The last two terms describe the reduction of the ground state population by absorption of signal and pump waves, respectively. In the steady-state, the ground state population is

$$n_0 = \frac{1}{\omega_c}\left(\frac{\rho}{T_1} + \frac{\beta_s}{I_s}\right) \tag{3.4}$$

where the "CPO center frequency" ω_c has been defined as

$$\omega_c = \frac{1}{T_1} + \frac{\alpha_p I_p}{\rho} + \frac{(\alpha_s + \beta_s) I_s}{\rho}. \tag{3.5}$$

This frequency will later be shown to determine the width of the spectral hole. Note that it is the inverse of the excited state lifetime, but also contains power-broadening terms for both pump and signal fields.

The signal field can be expressed as $I_s(t) = I_0 + I_1 \cos \delta t$, with a constant background component I_0 and a modulated portion I_1 oscillating at frequency δ. The ground state population can similarly be expressed as $n(t) = n_0 + n_1(t)$, with the same steady-state portion as before and an oscillating component $n_1(t)$. Plugging these expressions for $n(t)$ and $I_s(t)$ into (3.3), we get a linear differential equation for

the oscillating component $n_1(t)$ which can be quickly solved to find

$$n_1(t) = \left(\frac{\omega_c \cos \delta t + \delta \sin \delta t}{\omega_c^2 + \Delta^2} \right) I_1 g, \tag{3.6}$$

with a gain coefficient g defined to be

$$g = -\frac{n_0}{\rho}(\alpha_s + \beta_s) + \beta_s. \tag{3.7}$$

We construct the propagation equation for $I_s(t)$ by considering its z-dependence. As the signal wave propagates, it will lose intensity to absorption and gain intensity from stimulated emission. In this rate equation formulation, there are no coherent interactions between pump and signal waves. We also ignore scattering and other sources of loss in the fiber, as they are negligible compared to the near-resonance effects. Thus, we arrive at the propagation equation given in [31, 80],

$$\frac{dI_s}{dz} = -\frac{n}{\rho}\alpha_s I_s + \left(1 - \frac{n}{\rho}\right) \beta_s I_s. \tag{3.8}$$

To determine the behavior of I_1, we substitute our expressions for $n(t)$ and $I_s(t) = I_0 + I_1 \cos(\delta t + \phi)$, where we have allowed for a modulation phase shift ϕ during propagation, into (3.8). Isolating the terms that oscillate at the modulation frequency and ignoring terms of second-order or higher in I_1, we have

$$\frac{dI_1}{dz} = g I_1 \left(1 - \left(\frac{\alpha_s + \beta_s}{\rho}\right) \left(\frac{\omega_c}{\omega_c^2 + \delta^2}\right) I_0\right) \tag{3.9a}$$

$$\frac{d\phi}{dz} = \left(\frac{\alpha_s + \beta_s}{\rho}\right) \left(\frac{\delta}{\omega_c^2 + \delta^2}\right) I_0 g \tag{3.9b}$$

We can define a modulation absorption coefficient α_1 and modulation phase shift

ϕ_1 as

$$\alpha_1(I_1, I_p, I_0) = \left(\frac{\alpha_s + \beta_s}{\rho}\right)\left(\frac{\omega_c}{\omega_c^2 + \delta^2}\right) I_1 g \tag{3.10a}$$

$$\phi_1(I_1, I_p, I_0) = \left(\frac{\alpha_s + \beta_s}{\rho}\right)\left(\frac{\delta}{\omega_c^2 + \delta^2}\right) I_1 g \tag{3.10b}$$

which gives a simpler form of equations (3.9),

$$\frac{dI_1}{dz} = gI_1 - \alpha_1 I_0 \tag{3.11a}$$

$$\frac{d\phi}{dz} = \frac{I_0}{I_1}\phi_1 \tag{3.11b}$$

The evolution of the amplitude of the modulation depends on two terms, the first being a simple linear gain process and the second a modulation gain that depends on the carrier wave I_0. Note that since α_1 shares the sign of g, it will always be a modulation absorption effect if the system is inverted ($g > 0$), and a modulation gain effect if the system is lossy ($g < 0$). From this it is clear that this effect is always a spectral hole in the gain or absorption seen by I_1.

The evolution of the phase of the modulation has only one term which depends on I_0, I_1, and the phase shift ϕ_1. ϕ_1 shares the sign of α_1 and g, indicating that in a region of gain ϕ_1 will be positive and the modulation will experience advancement; conversely, in a region of loss ϕ_1 will be negative and the modulation will be delayed. From the derivation it is clear that α_1 is taken from the in-phase part of n_1, while the phase shift ϕ_1 is caused by the out-of-phase term $\delta \sin \delta t$.

Further, one can see from ϕ_1 that the maximum value of the phase shift should occur when the modulation frequency δ matches the CPO center frequency ω_c. The system can therefore be tuned for a particular modulation frequency by optimizing ω_c through power broadening. Conversely, at a given pump power or gain level, the modulation frequency can be adjusted until maximum advancement is achieved. The ideal modulation frequency will be on the order of $1/T_1$, which for un-pumped erbium

ions (T_1 = 10.5 ms) is approximately 100 Hz. Power broadening generally increases this frequency by approximately one order of magnitude.

3.3 Experimental procedure and results

A schematic diagram of our experimental setup is shown in Figure 3.3. A continuous-wave 1550-nm tunable diode laser (New Focus Velocity TLB-6328) serves as our signal beam. The signal beam is intensity-modulated by a free-space electro-optic modulator (New Focus 4104) to create approximately 0.5-ms full-width at half maximum pulses or 1 kHz sinusoidal intensity modulations. A sharp optical triggering signal which is too short to experience the CPO effect was added to each waveform for synchronization and triggering purposes. In both cases the signal modulation is superposed on a constant background intensity approximately ten times the size of the modulation depth to help reduce distortion effects.

The signal is then fiber-coupled and sent through an optical isolator, and a splitter sends a portion of the beam to an InGaAs photodiode for use as a reference. The remaining signal is combined with a continuous-wave 980-nm pump beam in a wavelength-division multiplexer (WDM). The combined pump and signal are then sent to one of two experimental configurations. Pump and signal powers at this point in the system were approximately 128 mW and 0.5 mW, respectively.

In the first configuration, shown in Figure 3.3(b), the pump and signal co-propagate through a series of three 3-m sections of EDOF separated by bi-directional 1% fiber-optic couplers. After passing through all nine meters of EDOF, the signal is separated from the pump by a WDM and detected with a photodiode. The same WDM and photodiode were then used to measure the four outputs of the bi-directional taps labeled A through D to determine direction of energy flow and approximate signal strength. Signal and reference photodiode outputs were monitored and captured with a digital storage oscilloscope.

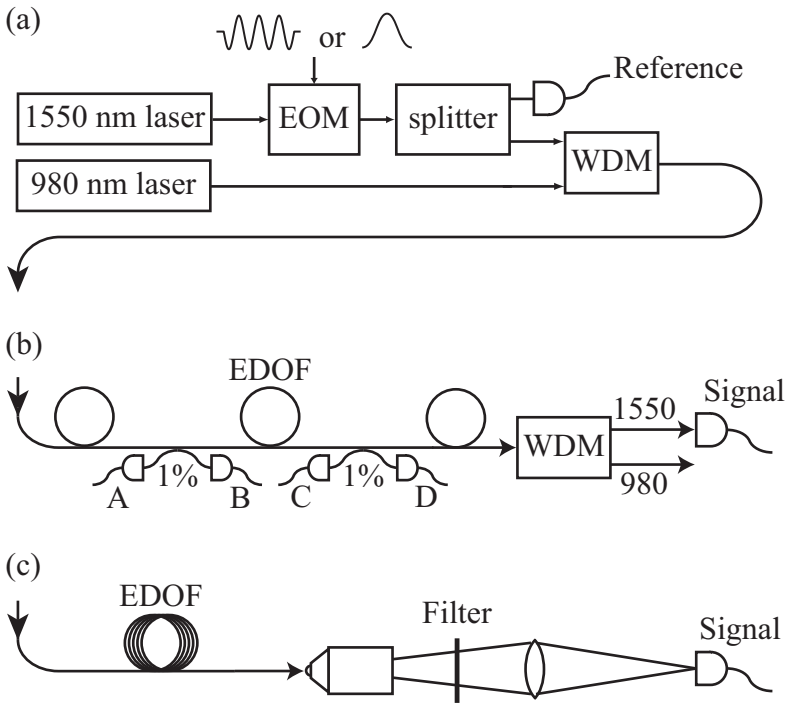

Figure 3.3: Experimental setup (a)-(c) for the backwards propagation experiment. Abbreviations used: EOM = electro-optic modulator, WDM = wavelength division multiplexer, EDOF = erbium doped optical fiber. Setup (a) acted as the source for all experiments. Setup (b) was used to measure the field evolution within the fiber, while setup (c) was used to determine energy transport dynamics.

As expected, measurements made on ports B and D showed strong signals for both modulation formats, as did the output of the full nine meters of EDOF. In all cases, the modulation was advanced and slightly distorted by the saturated amplification process. The total advancement measured was approximately 120 μs for the full 9 meters of fiber, corresponding to a velocity of -75 km/s and a group index of $n_g \approx -4000$.

However, measurements on ports A and C did not show an appreciable signal pulse. The intensity at these two ports was barely distinguishable from the amplified spontaneous emission (ASE) noise measured in the absence of a signal pulse. The largest feature of these measurements was a *dip* in ASE background corresponding

to the temporary saturation of the gain during the pulse transit. In addition, we observed a small amount of energy due to back-reflections from the numerous fiber splices in the system, though this was negligible compared to the ASE background or signal powers.

These results confirmed that the energy flow in this system was confined to the forward direction despite the appearance of pulse advancement. This is an important conclusion for several reasons. It supports the proposed theoretical treatments of negative-group-velocity materials, which suggest that there is no energy traveling in the backward direction. It also demonstrates that the energy velocity and group velocity can not only differ in magnitude, but that they can differ in sign as well. Most importantly, it confirms that the behavior of the system at any point in the fiber is not dependent on points further along in the fiber. This justification allows us to continue with the second stage of our experiment and investigate the time evolution of the pulse within the EDOF.

In this final stage of our experiment, we send the signal and pump from Figure 3.3(a) through the setup shown in Figure 3.3(c). The two beams propagate through a nine-meter coil of EDOF, after which they are coupled into free space with a microscope objective and index-matching fluid. A bandpass filter isolated the signal at 1550 nm, which was imaged to a germanium photodetector for measurement. The signal and reference detector outputs were again recorded with the digital storage oscilloscope for both modulation formats.

The fiber was then cut with a mechanical cleaving tool to reduce its length by approximately 25 cm and re-coupled to the objective. New measurements were taken with this new fiber length, and the process was repeated until only a few centimeters of fiber remained. In this fashion, we were able to collect data for a variety of fiber lengths and reconstruct the time evolution of the pulse at many points along the length of the fiber.

Example signal and reference pulses for propagation through six meters of EDOF

are shown in Figure 3.4. The signal pulse has been amplified considerably, so the intensities have been normalized to facilitate comparison. The output pulse is noticeably advanced compared to the input pulse, and a small amount of pulse distortion is observed.

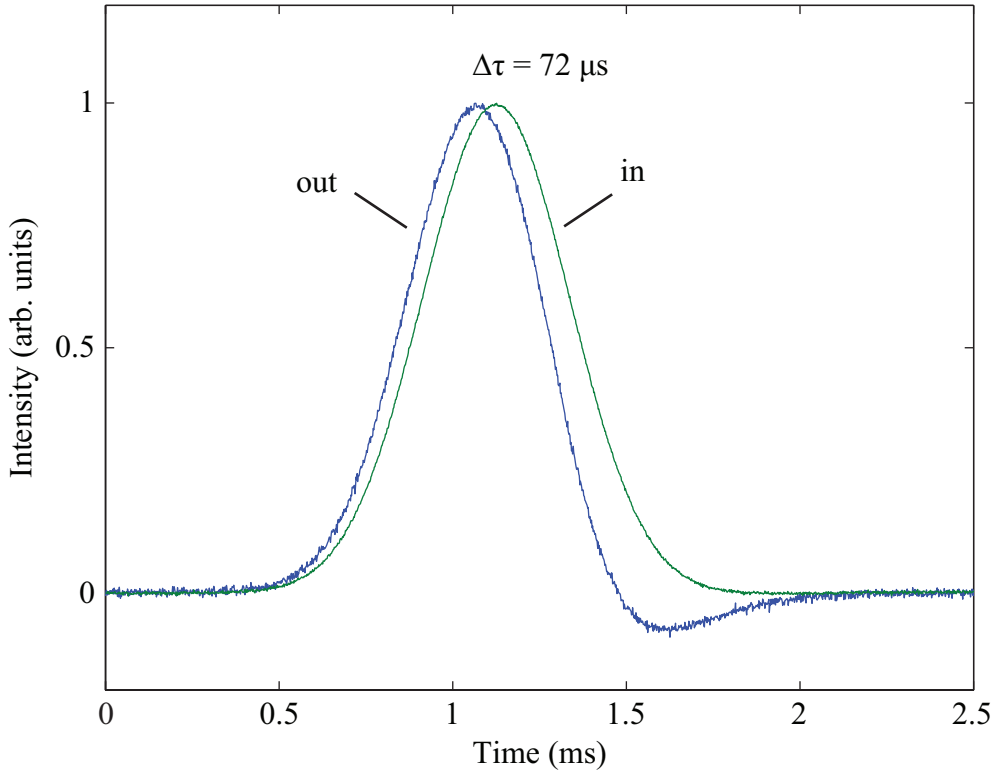

Figure 3.4: Pulse intensity profiles for the input (green) and output (blue) of 6 meters of erbium-doped optical fiber. Pulse intensities have been normalized to clarify the comparison of input and output pulse shapes. A constant background power equal to approximately ten times the pulse modulation amplitude has been suppressed in normalization. The output pulse is advanced in time by 72 μs and suffers a slight amount of distortion.

Our data collection gives us a series of temporal intensity profiles $I(t)$ at different lengths L_i of EDOF, which can be treated as a two-dimensional map of $I(z,t)$, the spatial and temporal intensity profile. Each individual trace $I(t)$ is simply $I(L_i,t)$ evaluated at a fixed value of $z = L_i$. If we instead fix the value of t at a given time step $t = t_i$, we have $I(z) = I(z,t_i)$, the spatial intensity profile within the fiber. This

technique is validated by our first measurement, which confirmed that downstream portions of the fiber have no effect on the energy or pulse propagation dynamics further upstream.

Figure 3.5 illustrates the result, showing snapshots of normalized pulse intensity $I(z, t_i)$ as a function of normalized fiber position $|n_g|z$ for a series of time steps. Pulse intensity has been normalized after background subtraction according to the maximum value of $I(L_i, t)$ at each fiber length L_i in the same fashion as Figure 3.4. The small noise-like intensity fluctuations in the fiber region are a result of changes in coupling and detection efficiency in individual traces, though these effects have been partially mitigated by the normalization scheme. The group index has been chosen as $n_g = -4000$ for these plots based on our advancement measurements. The green vertical bars indicate the entrance face of the fiber at $|n_g|z = 0$ km and the exit face at $|n_g|z = 36$ km. Data points outside the fiber region are interpolated from the measurements made at $L = 0$ m and $L = 9$ m assuming normal propagation at $v_g = c/n$. Arrows have been provided to help identify the approximate position of the peak in all three regions.

It's clear that the pulse peak exiting the fiber is created before the second snapshot even though the peak of the incident pulse does not arrive at the entrance face of the fiber until the last time step. Between these two times, a peak appears to be created within the fiber that moves in the backward direction (from right to left) linking input and output pulse peaks. The apparent "backwards propagation" occurs because of time-dependent transfer of energy between the optical pulse and the gain medium, which reshapes the pulse profile during propagation. The peak of the optical intensity within the fiber coincides with a minimum of the stored energy density in the material, as predicted theoretically [24, 81]; this is further substantiated by the observation of a minimum in the amplified spontaneous emission power in the energy measurements made on ports A and C of the first experiment.

While the pulse envelope does indeed appear to propagate backwards within the

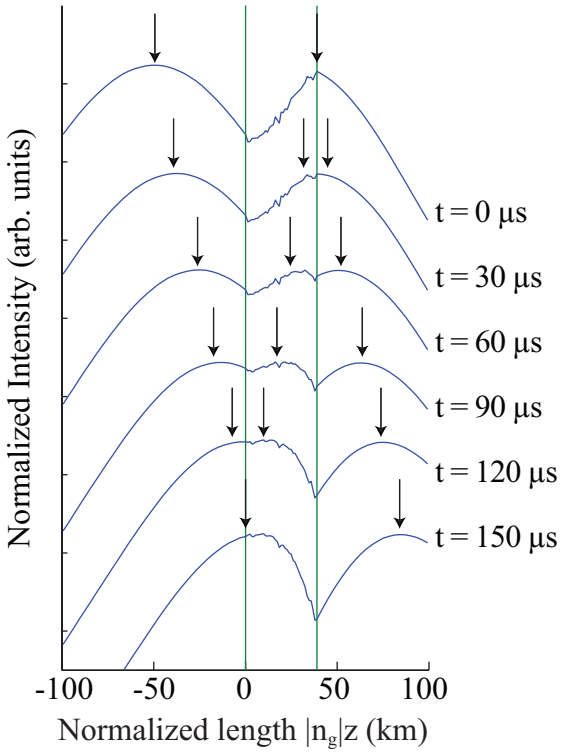

Figure 3.5: Snapshots of normalized pulse intensity $I(z,t)$ showing the time evolution of the optical intensity before (left), after (right), and within the erbium-doped fiber (center). The edges of the EDOF are denoted by the vertical green lines, and the times are given in microseconds. The group index for the erbium fiber region is approximately $n_g = -4000$. Arrows mark the approximate position of the peak.

fiber, the energy flow is always in the forward direction, suggesting that there is no physically meaningful object moving backwards in the system. As such, there is also no violation of causality in this system, since the peak exiting the material is created from the leading edge of the input pulse as a result of gain saturation in the EDOF. Similarly, the incident pulse peak becomes part of the tail of the exiting pulse, confirming that the incident and exiting peaks are not causally related objects.

To demonstrate that the backwards propagation effect is not a side effect of the normalization procedure, we provide the un-normalized data set in Figure 3.6. The gain of the EDOF causes the output pulse to be much larger than the input pulse, so

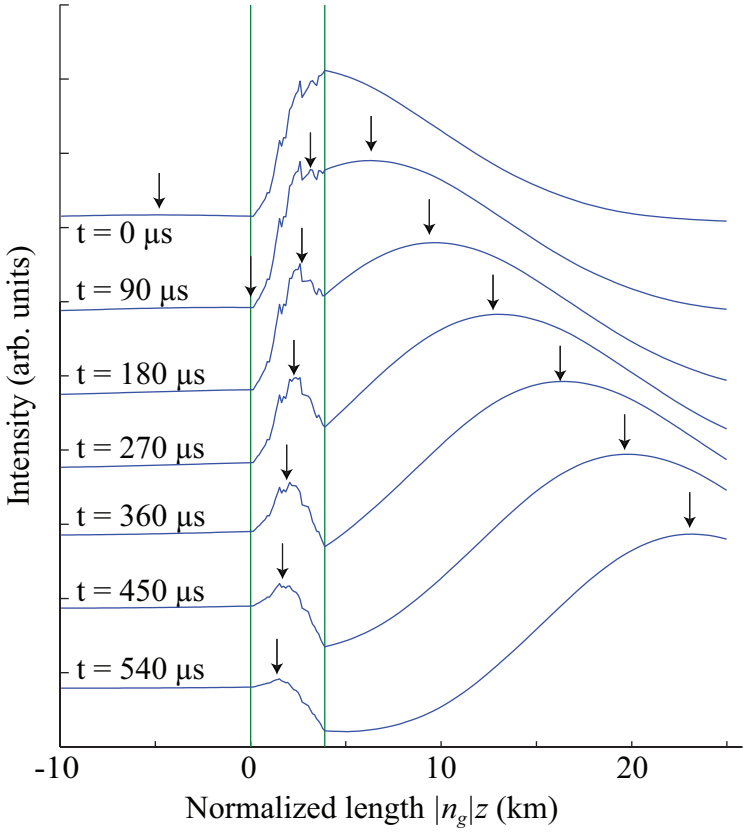

Figure 3.6: Snapshots of unnormalized pulse intensity $I(z,t)$ at a series of time steps. The output pulse is much larger than the input pulse, making it difficult to see the input peak. Arrows have been provided to help identify the peak positions more clearly. The pulse peak within the material still appears to move backwards, but has a different apparent group velocity as a consequence of the influence of gain saturation effects.

arrows have been provided to help identify the peak positions. The noise-like intensity fluctuations caused by coupling efficiency changes are more pronounced in the absence of the normalization procedure. In this data set, the pulse still appears to move backwards through the EDOF section, but the apparent group velocity is considerably smaller and the input peak no longer appears to coincide with the backward-propagating peak at the first interface. This strange effect is a result of the non-uniformity of the longitudinal gain distribution within the fiber.

While a uniform longitudinal gain profile $g(z) = g_0$ would not affect the expected

group velocity of a Gaussian pulse within the material, $g(z)$ is neither uniform nor stationary in our system. The pump beam saturates the gain in the early portions of the fiber, but pump absorption leads to a reduction in the gain coefficient at longer fiber lengths. As such, the apparent group velocity observed in the un-normalized data is significantly different than the normalized data set. The nonlinear pulse propagation dynamics which saturate the gain and cause temporal variations in the longitudinal gain profile are *not* responsible for this change in group velocity, as those are not eliminated by the normalization process.

It should also be noted that after $t = 90$ μs the entire pulse envelope within the fiber is monotonically decreasing in intensity. The peak intensity at $z = 0$ occurs at the same time on either side of the interface, even though a point of higher intensity exists further along in the fiber. The pulse peak in the material is thus a result of the complicated time-dependent longitudinal gain profile in the fiber, which is affected by gain saturation in the EDOF due to the signal as well as gain recovery caused by pump and signal absorption. In the limit of small gain ($g_0 L < 1$), the effects of gain non-uniformity become very weak and the un-normalized data behaves exactly like the normalized version shown in Figure 3.5.

It is for this reason that we feel the normalized data gives us better intuition for the behavior within the fiber. It accurately describes the time at which the optical intensity $I(z_i, t)$ reaches its maximum value, which is a stronger indicator that the material gain at z_i has been fully saturated. The normalization procedure removes the complicating effects of gain non-uniformity caused by pump depletion while retaining the dynamic nonlinear gain saturation effects that create the fast-light effect and cause pulse advancement.

3.4 Summary

In this chapter, we introduced the concept of negative group velocity and reviewed the predictions that have been made about pulse propagation under those conditions. These predictions were demonstrated with a simple simulation result, and the limitations of both the simulation and practical negative-group-velocity materials were discussed.

We also presented a brief theoretical treatment of coherent population oscillations in erbium-doped optical fiber systems as a mechanism to achieve slow- and fast-light. The theory demonstrated that it is capable of providing either pulse delay or pulse advancement for the 1550-nm signal wavelength contingent on the absence or presence of a pump laser at the 980-nm transition.

Finally, we described the first experimental measurements of pulse evolution and energy dynamics within an EDOF operating in the negative-group-velocity regime. We have found that the energy velocity is always positive in such a system, consistent with theoretical predictions. In addition, we have observed the presence of a backward-propagating peak within the material that links the input and output pulses.

These results can be understood by properly including the time-dependent saturation of the gain material in our conceptual model. The leading edge of the incident pulse experiences a large amount of gain, but at some point the gain becomes saturated and a new peak is created at the exit interface. This gain saturation effect starts at the longest point in the fiber and slowly propagates back through the fiber towards the entrance face as the input intensity increases further. This subsequently causes a local maximum of pulse intensity within the fiber at the vicinity of this transient saturation point, resulting in the appearance of a backward-propagating intensity maximum that we identify as a peak. In systems where the gain is severely non-uniform, which is frequently the case due to pump depletion, the intensity profile in the fiber will be distorted but the propagation dynamics can be recovered by an

appropriate normalization technique.

Chapter 4

Tunneling delays in one dimension

The concept of tunneling through a potential barrier is usually encountered very early in a student's undergraduate career. The finite barrier problem in particular is frequently used to demonstrate the differences between classical and quantum-mechanical behavior. However, within this simple undergraduate problem lies a deeper and more intriguing question about the time associated with the tunneling process. Complicating matters further, there is some uncertainty about the appropriate way to calculate such a tunneling time and no universally agreed-upon interpretation of the physical meaning of this time.

In this chapter, we will provide a complete derivation of the tunneling delay from first principles, and express this quantity using two different interpretations. The derivation is heavily based on papers by Steinberg [82] and Winful [83], primarily using the notation of the latter. There are a number of other definitions of tunneling time that we will mention in passing [84, 85].

4.1 The one-dimensional finite barrier problem

Figure 4.1 shows a standard one-dimensional finite potential barrier problem. A particle of energy E in a region of potential energy $V = 0$ is incident upon a barrier

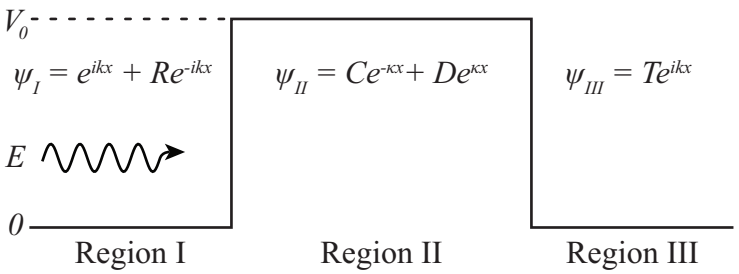

Figure 4.1: Schematic diagram for one-dimensional barrier tunneling. A particle of energy E approaches a barrier of potential V_0. ψ is the wavefunction in each region.

region of potential energy $V_0 > E$ at $x = 0$. The barrier extends from $x = 0$ to $x = L$, after which the potential energy returns to $V = 0$. To describe the particle wavefunction in each region, we consult the time-dependent Schrödinger equation,

$$\left[\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} - V(x)\right]\Psi(x,t) = -i\hbar\frac{\partial}{\partial t}\Psi(x,t). \tag{4.1}$$

From this, we see that the wavefunction has solutions of the form

$$\Psi(x,t) = \Psi_E(x)e^{-iEt/\hbar}, \tag{4.2}$$

where $\Psi_E(x)$ is the stationary state solution to the time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi_E}{\partial x^2} + (V - E)\Psi_E = 0. \tag{4.3}$$

In regions I and III, $V = 0$ and Ψ_E has traveling-wave solutions of the form $e^{\pm ikx}$, with $E = \hbar^2k^2/2m$. Inside the barrier, the wave is evanescent, and solutions take the form $e^{\pm\kappa x}$, with $\kappa \equiv \sqrt{2m(V_0 - E)}/\hbar$. Thus, the stationary state wavefunctions in

regions I-III can be expressed as

$$\Psi_{\mathrm{I}}(x) = e^{ikx} + Re^{-ikx} \tag{4.4a}$$

$$\Psi_{\mathrm{II}}(x) = Ce^{-\kappa x} + De^{\kappa x} \tag{4.4b}$$

$$\Psi_{\mathrm{III}}(x) = Te^{ikx} \tag{4.4c}$$

where we have assumed that there are no particles incident from the right ($x = \infty$). Finding T, R, C, and D is a straightforward, if tedious, exercise in algebra. At $x = 0$, both Ψ and $\frac{\partial \Psi}{\partial x}$ must be continuous,

$$\Psi_{\mathrm{I}}(0) = \Psi_{\mathrm{II}}(0) \tag{4.5a}$$

$$\Psi_{\mathrm{I}}^{'}(0) = \Psi_{\mathrm{II}}^{'}(0) \tag{4.5b}$$

Plugging (4.4) into (4.5) gives

$$1 + R = C + D \tag{4.6a}$$

$$ik(1 - R) = -\kappa(C - D) \tag{4.6b}$$

from these two equations, we can express C, D, and $\Psi_{II}(x)$ in terms of R:

$$2C = 1 + R - \frac{ik}{\kappa}(1 - R) \tag{4.7a}$$

$$2D = 1 + R + \frac{ik}{\kappa}(1 - R) \tag{4.7b}$$

$$\begin{aligned} \Psi_{\mathrm{II}}(x) &= (1 + R)\left(\frac{e^{\kappa x} + e^{-\kappa x}}{2}\right) + \frac{ik}{\kappa}(1 - R)\left(\frac{e^{\kappa x} + e^{-\kappa x}}{2}\right) \\ &= (1 + R)\cosh \kappa x + \frac{ik}{\kappa}\sinh \kappa x \end{aligned} \tag{4.7c}$$

Next, we'll apply boundary conditions at $x = L$:

$$\Psi_{\mathrm{II}}(L) = \Psi_{\mathrm{III}}(L) \tag{4.8a}$$

$$\Psi'_{\mathrm{II}}(L) = \Psi'_{\mathrm{III}}(L) \tag{4.8b}$$

Plugging in (4.4) and (4.7c) gives

$$(1+R)\cosh\kappa L + \frac{ik}{\kappa}(1-R)\sinh\kappa L = Te^{ikL} \tag{4.9a}$$

$$\kappa(1+R)\sinh\kappa L + ik(1-R)\cosh\kappa L = ikTe^{ikL} \tag{4.9b}$$

which can be used to solve for T and R. Since eliminating T is easy, we solve R first. Dividing (4.9b) by ik and setting the left-hand sides of (4.9) equal gives:

$$(1+R)\cosh\kappa L + \frac{ik}{\kappa}(1-R)\sinh\kappa L = \frac{\kappa}{ik}(1+R)\sinh\kappa L + (1-R)\cosh\kappa L$$

$$2R\cosh\kappa L + i\left[\frac{k}{\kappa}(1-R) + \frac{\kappa}{k}(1+R)\right]\sinh\kappa L = 0$$

$$2R\cosh\kappa L + i\left(\frac{\kappa}{k} - \frac{k}{\kappa}\right)R\sinh\kappa L = -i\left(\frac{k}{\kappa} + \frac{\kappa}{k}\right)\sinh\kappa L$$

For convenience, we define two parameters that simplify the notation:

$$\Delta \equiv \frac{1}{2}\left(\frac{\kappa}{k} - \frac{k}{\kappa}\right) \tag{4.10a}$$

$$\Delta' \equiv \frac{1}{2}\left(\frac{\kappa}{k} + \frac{k}{\kappa}\right) \tag{4.10b}$$

with these, R becomes

$$R = \frac{-i\Delta'\sinh\kappa L}{\cosh\kappa L + i\Delta\sinh\kappa L}.$$

Here we define one more parameter to simplify the notation. Following convention [86], we define g such that

$$g \equiv \cosh \kappa L + i\Delta \sinh \kappa L, \tag{4.11}$$

which then gives R a very simple form:

$$R = -\frac{i\Delta'}{g} \sinh \kappa L \tag{4.12}$$

Solving for T is easiest by evaluating (4.9a)+$\frac{1}{ik}$(4.9b) and plugging in (4.12):

$$\begin{aligned}
2Te^{ikL} &= (1+R)\cosh \kappa L + \frac{ik}{\kappa}(1-R)\sinh \kappa L + \frac{\kappa}{ik}(1+R)\sinh \kappa L + (1-R)\cosh \kappa L \\
&= 2\cosh \kappa L + i\left(\frac{k}{\kappa} - \frac{\kappa}{k}\right)\sinh \kappa L - iR\left(\frac{\kappa}{k} + \frac{k}{\kappa}\right)\sinh \kappa L \\
&= 2(\cosh \kappa L - i\Delta \sinh \kappa L - iR\Delta' \sinh \kappa L) \\
&= \frac{2}{g}\left[g(\cosh \kappa L - i\Delta \sinh \kappa L) - \Delta'^2 \sinh \kappa L^2\right] \\
&= \frac{2}{g}\left[\cosh \kappa L^2 + \Delta^2 \sinh \kappa L^2 - \Delta'^2 \sinh \kappa L^2\right] \\
&= \frac{2}{g}\left[\cosh \kappa L^2 + (\Delta^2 - \Delta'^2) \sinh \kappa L^2\right] \\
&= \frac{2}{g}\left[\cosh \kappa L^2 - \sinh \kappa L^2\right] = \frac{1}{g} \\
T &= \frac{e^{-ikL}}{g}
\end{aligned} \tag{4.13}$$

Finding C and D is also straightforward from (4.12) and (4.7). By noting that $1 \pm R = (g \mp i\Delta' \sinh \kappa L)/g$, $\Delta + \Delta' = \kappa/k$, and $\Delta - \Delta' = -k/\kappa$, C quickly simplifies

to

$$\begin{aligned}
2C = (1+R) - \frac{ik}{\kappa}(1-R) &= \frac{1}{g}\left[g - i\Delta' \sinh\kappa L - \frac{ik}{\kappa}g + \frac{k}{\kappa}\sinh\kappa L\right] \\
&= \frac{1}{g}\left[\cosh\kappa L + i(\Delta - \Delta')\sinh\kappa L - \frac{ik}{\kappa}\cosh\kappa L + \frac{k}{\kappa}(\Delta + \Delta')\sinh\kappa L\right] \\
&= \frac{1}{g}\left[\cosh\kappa L + \sinh\kappa L - i\frac{k}{\kappa}(\cosh\kappa L + \sinh\kappa L)\right] \\
&= \left(1 - \frac{ik}{\kappa}\right)e^{\kappa L}/g \\
C &= \left(1 - \frac{ik}{\kappa}\right)e^{\kappa L}/2g \qquad (4.14)
\end{aligned}$$

In a similar fashion, D can be calculated as

$$D = \left(1 + \frac{ik}{\kappa}\right)e^{-\kappa L}/2g \qquad (4.15)$$

Equations (4.12), (4.13), (4.14) and (4.15), along with the definitions of g, Δ, and Δ' completely describe the one-dimensional system's stationary states. Figure 4.2 shows the probability distribution $P(x) = |\Psi(x)|^2$ for such a state with $\kappa = k/3$, which is equivalent to $E = 0.9V$. The incident and reflected parts of the wave function interfere to form a standing wave in region I, and the probability distribution decays as $e^{-\kappa x}$ in the barrier region. The portion of the wave function describing the tunneled particle in region III has the form of a traveling wave, and therefore has a constant probability density. Since a single stationary state wave function exists at all points in space and time, and does not have a moving "peak" that can be unambiguously identified, we cannot infer a useful definition of tunneling time from one such state.

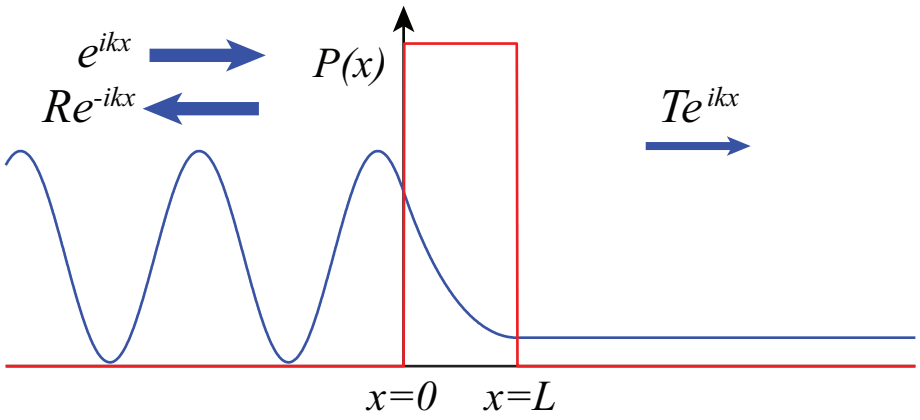

Figure 4.2: Probability density function for the stationary state tunneling process.

4.2 Wave packets and the phase time interpretation of group delay

To discuss group delay, we need to construct a spatially localized particle or wave packet. As in [83], we do this by summing over a narrow band of stationary states:

$$\Psi(x,t) = \int_E f(E - E_0)\Psi_{\mathrm{E}}(x)e^{-iEt/\hbar}dE, \tag{4.16}$$

where f is an energy distribution function that describes the energy components of the wave packet, and $\Psi_E(x)$ are the steady-state wavefunctions of energy E derived in the previous section. The wavefunction of the incident particle is found by substituting (4.2) for region I into (4.16). We can describe the motion of the particle by determining the time evolution of the peak, which exists where the phases of the wavefunction's constituent steady-states interfere constructively.

4.2.1 The phase time interpretation of group delay

Through the method of stationary phase, we can obtain a constraint equation for the time evolution of the peak,

$$\frac{\partial}{\partial E}\left(\arg\left(\Psi_{\mathrm{E}}\right) - Et/\hbar\right) = 0. \tag{4.17}$$

By our definitions of Ψ_E in (4.4a), this occurs at $x = 0$, $t = 0$ for the wave packet incident from the left in the absence of a barrier. This expression also describes the wave packet's propagation in region I far in front of the barrier, and can be evaluated to give the velocity of the particle,

$$\begin{aligned}
\frac{\partial k}{\partial E}x - t/\hbar = \frac{\sqrt{2m}}{\hbar}\frac{1}{2\sqrt{E}}\,x - \frac{t}{\hbar} &= 0 \\
x &= \sqrt{\frac{2E}{m}}\,t \\
&= \frac{\hbar k t}{m} \\
v = \frac{\partial x}{\partial t} &= \frac{\hbar k}{m}.
\end{aligned}$$

We can also interpret v as the incident particle flux in region I.

The transmitted wave packet is described by substituting (4.4c) into (4.16),

$$\psi_T(x,t) = \int_E f(E - E_0)|T(E)|e^{i\phi_T(E) + ikx - iEt/\hbar}dE. \tag{4.18}$$

If $|T(E)|$ does not vary greatly over the region where $f(E - E_0)$ is significant, the wave packet is not distorted or reshaped upon transmission. The peak "leaves" the exit face of the barrier under the same stationary phase condition, namely that

$$\frac{\partial}{\partial E}\left(\arg\left(\Psi_{III}(L)\right) - Et/\hbar\right) = 0 \tag{4.19}$$

where $\arg(\Psi_{III}(L)) = \arg(T) + kL$. If we say that this occurs at time $t = \tau_{gt}$, and let

$T = |T|e^{i\phi_T}$, we have

$$\tau_{gt} = \hbar \frac{\partial}{\partial E} \left(\phi_T + kL \right). \tag{4.20}$$

If $|R(E)|$ is also slowly-varying over the region of interest, we can define a delay time for the reflected wave packet in a similar fashion. Differentiating the phase of the reflected component far from the barrier, we have

$$\frac{\partial}{\partial E} \left(\arg(R) - Et/\hbar \right) = 0. \tag{4.21}$$

Evaluating this for $R = |R|e^{i\phi_R}$ gives us the delay of the reflected wave packet τ_{gr} in terms of ϕ_R,

$$\tau_{gr} = \hbar \frac{\partial \phi_R}{\partial E}. \tag{4.22}$$

The condition that $|T(E)|$ or $|R(E)|$ not vary greatly over the region where f is significant can be interpreted as limiting the scope of these expressions to the "quasi-static" regime. If this condition is not obeyed, the fluctuations in $|T(E)|$ can have a significant effect on the integral and the wave packet may begin to experience distortion or reshaping which can lead to pulse breakup and ambiguity in the output pulse's peak.

Put another way, these equations retain their relevance as long as the localization of the particle, or the particle "length," is much broader than the barrier length. If the particle length is reduced enough, it will contain energy components that lie outside the barrier region and transient effects begin to occur that distort the particle and the equations above lose their validity. In the extreme limit of a discontinuous particle wave function, the discontinuity will propagate at c through the barrier region, and the particle will exhibit extreme amounts of distortion.

The tunneling process for a spatially localized particle is illustrated schematically in Figure 4.3. In the first panel a particle is incident upon the barrier region from the left. As the peak approaches the left side of the barrier, the incident and reflected portions of the particle wave function begin to interfere just as in the stationary state

case of Figure 4.2. The visibility of this interference becomes maximum at $t = 0$, the time at which the peak of the incident particle would have crossed $x = 0$ in the absence of the barrier structure. The standing wave that this creates is a transient effect which lasts only for the duration of the pulse overlap. Within the barrier the particle wave function is decaying as $e^{-\kappa x}$ as it undergoes the tunneling process, with its amplitude adiabatically following the amplitude of the standing wave at $x = 0$. In the last panel the incident particle has been reflected or transmitted, and the transmitted particle's peak exhibits advancement relative to a copy of the incident particle that is scaled by the barrier transmission coefficient but does not otherwise experience the barrier, represented by a green dashed line.

The interference of incident and reflected waves makes it difficult to accurately determine an arrival time for the incident wave packet at the barrier interface. As such, we extrapolate this arrival time from the behavior of the incident wave packet far from the barrier to determine the time at which it *should* have arrived at $x = 0$ had the barrier not been present. The departure time of the reflected particle is similarly determined based on its behavior long after the tunneling process has taken place. It is therefore important to recognize that the group delays in equations (4.20) and (4.22) are quantities that describe completed tunneling events based on extrapolation from measurements made far from the barrier, as shown in the third panel of Figure 4.3.

It is a common mistake to misinterpret these group delays as traversal or propagation times, raising questions about causality violation [4–7]. The error in this logic is the false assumption that the peaks of the incident and transmitted waves are causally-related entities. These peaks are the result of interference between many stationary states, and do not represent a localized physical object propagating from entrance to exit. As Winful states in [87], "We cannot say where the transmitted wave packet is at $t = 0$ and hence cannot say that the group delay measures the time it takes a wave packet to travel from input to output." Furthermore, the wave func-

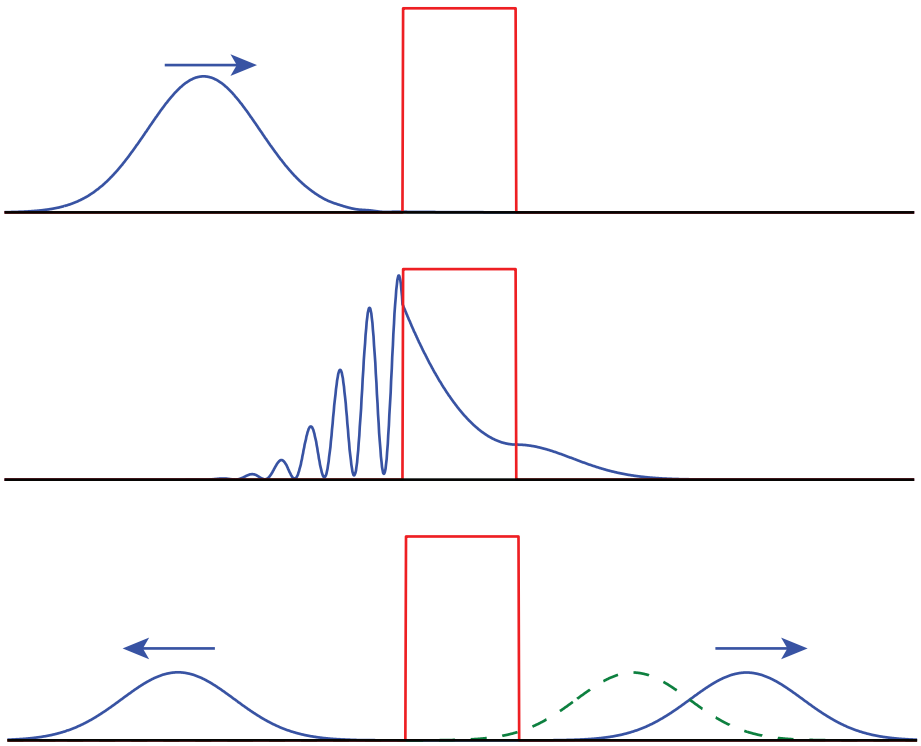

Figure 4.3: Schematic showing the tunneling dynamics of a particle composed of many stationary states. During the tunneling process, the incident and reflected portions of the particle wave function interfere to form a standing wave. The tunneled portion decays in the barrier region, and the transmitted particle exhibits advancement relative to a scaled copy of the incident wave that has been causally propagated through an equivalent length of free space, shown in green.

tion probability density in the transmitted region never exceeds that of a "reference" version of the incident wave function that has been causally propagated [88].

This interpretation has prompted arguments that the group delay itself is not physically meaningful [89]. On the contrary, it has a very clear physical significance: it accurately describes the behavior of the peak of the transmitted and reflected particle wavefunctions after experiencing the tunneling process [90]. It is the misinterpretation of this quantity as a traversal time that lacks physical meaning. One would not claim to measure the speed of a train by comparing the time at which the front of the train exits a tunnel to the time at which the back of the train enters it, because the front

and back of the train are distinct and different physical objects. The same is true in the case of tunneled particles, because the peaks of the incident and tunneled wave functions are not the same entity.

4.2.2 Reflection and transmission delay for asymmetric and symmetric barriers

For a symmetric barrier $\tau_{gt} = \tau_{gr}$ due to time-reversal symmetry and particle conservation. This is discussed in [91] and demonstrated in the appendix of [92] and in greater detail in Section III and Appendix A of [93]. Here we provide a consolidated derivation based on these sources using the system shown in Figure 4.4, which is an asymmetric barrier with an arbitrary potential $V_{\mathrm{II}} = V(x)$ in the barrier region and constant potentials $V_{\mathrm{I}} \neq V_{\mathrm{III}}$ and particle wavevectors k and $\tilde{k}$ in regions I and III respectively

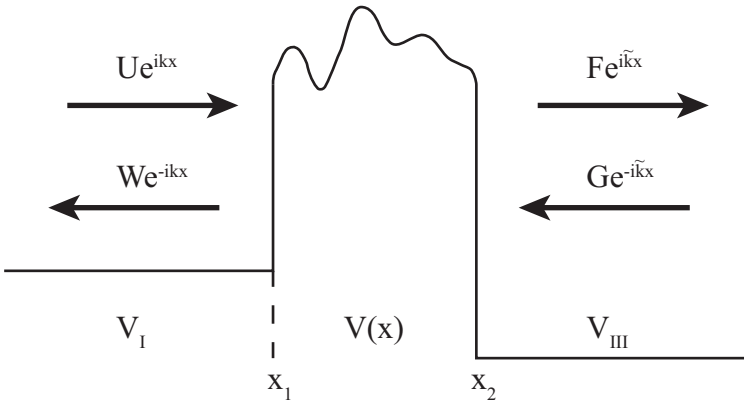

Figure 4.4: Asymmetric barrier system for calculation of τ_{gt} and τ_{gr} as shown in [93]. The potential energies are V_{I} , $V(x)$, and V_{III} in regions I-III respectively.

For this system, we can write a transfer matrix describing the transformation from the wavefunction $\Psi_2 = Fe^{i\tilde{k}x} + Ge^{-i\tilde{k}x}$ at point x_2 in region III to wavefunction

$\Psi_1 = Ue^{ikx} + We^{-ikx}$ at point x_1 in region I:

$$\begin{pmatrix} U \\ W \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} F \\ G \end{pmatrix} = M \begin{pmatrix} F \\ G \end{pmatrix} \tag{4.23}$$

The complex conjugate Ψ^* obeys the same Schrödinger equation as Ψ; therefore we can also write

$$\begin{pmatrix} W^* \\ U^* \end{pmatrix} = M \begin{pmatrix} G^* \\ F^* \end{pmatrix} \tag{4.24}$$

Examining this relation for the two cases shows that

$$U = M_{11}F + M_{12}G, \tag{4.25a}$$

$$W = M_{21}F + M_{22}G \tag{4.25b}$$

$$U^* = M_{21}G^* + M_{22}F^* \tag{4.25c}$$

$$W^* = M_{11}G^* + M_{12}F^* \tag{4.25d}$$

From the expressions for W and W^* we have $W = M_{12}^*F + M_{11}^*G = M_{21}F + M_{22}G$, implying that

$$M_{22} = M_{11}^*, \quad M_{21} = M_{12}^* \tag{4.26}$$

which is a statement of time-reversal symmetry.

Particle (or current) conservation gives us an additional constraint, namely that the particle current $\overline{J} = (\Psi^*\nabla\Psi - \Psi\nabla\Psi^*)$ is a conserved quantity [94]. Evaluation of this expression at points x_1 and x_2 gives us

$$\left(\Psi_1^*\frac{\partial\Psi_1}{\partial x} - \Psi_1\frac{\partial\Psi_1^*}{\partial x}\right) = \left(\Psi_2^*\frac{\partial\Psi_2}{\partial x} - \Psi_2\frac{\partial\Psi_2^*}{\partial x}\right) \tag{4.27}$$

Substitution of Ψ_1 into the left-hand side and simplifying gives

$$\left(U^*e^{-ikx}+W^*e^{ikx}\right)\frac{\partial}{\partial x}\left(Ue^{ikx}+We^{-ikx}\right)-\left(Ue^{ikx}+We^{-ikx}\right)\frac{\partial}{\partial x}\left(U^*e^{-ikx}+W^*e^{ikx}\right)$$
$$=2ik\left(|U|^2-|W|^2\right)$$

Similarly, substitution of Ψ_2 into the right-hand side gives

$$i\tilde{k}\left(F^*e^{-i\tilde{k}x}+G^*e^{i\tilde{k}x}\right)\left(Fe^{i\tilde{k}x}-Ge^{-i\tilde{k}x}\right)+i\tilde{k}\left(Fe^{i\tilde{k}x}+Ge^{-i\tilde{k}x}\right)\left(F^*e^{-i\tilde{k}x}-i\tilde{k}G^*e^{i\tilde{k}x}\right)$$
$$=2i\tilde{k}\left(|F|^2-|G|^2\right)$$

Which gives (4.27) the form

$$k\left(|U|^2-|W|^2\right)=\tilde{k}\left(|F|^2-|G|^2\right). \tag{4.28}$$

If we substitute (4.25a)-(4.25b) into this, we find that

$$\begin{aligned} k\left(M_{11}M_{22}-M_{12}M_{21}\right)\left(|F|^2-|G|^2\right)&=\tilde{k}\left(|F|^2-|G|^2\right)\\ \left(M_{11}M_{22}-M_{12}M_{21}\right)&=\tilde{k}/k\\ \det M&=\tilde{k}/k \end{aligned} \tag{4.29}$$

To derive the relation between reflection delay τ_{gr} and transmission delay τ_{gt}, we consider two situations. The first is the "direct" process of a particle incident on the barrier from the left; in this case $U=1$, $W=B$, $F=A$, and $G=0$, such that

$$\begin{pmatrix}1\\B\end{pmatrix}=M\begin{pmatrix}A\\0\end{pmatrix}$$

giving us the relations

$$M_{11}A = 1 \tag{4.30a}$$

$$M_{21}A = B \tag{4.30b}$$

and by inspection, $A = 1/M_{11}$. The "indirect" or reverse process is a particle incident from the right, for which $U = 0$, $W = \overline{A}$, $F = \overline{B}$, and $G = 1$. This gives us

$$\begin{pmatrix} 0 \\ \overline{A} \end{pmatrix} = M \begin{pmatrix} \overline{B} \\ 1 \end{pmatrix}$$

$$M_{11}\overline{B} + M_{12} = 0 \tag{4.31}$$

$$M_{21}\overline{B} + M_{22} = \overline{A} \tag{4.32}$$

From which it's clear that

$$\overline{A} = \frac{\det M}{M_{11}} = \frac{\tilde{k}}{k}A \tag{4.33}$$

and since k and $\tilde{k}$ are both real, the phases of $\overline{A}$ and A can be related simply by

$$\overline{\alpha} = \arg\left(\overline{A}\right) = \arg\left(A\right) = \alpha \tag{4.34}$$

This result is significant, because the group delay in transmission for the "direct" process is $\tau_{gt} = \hbar\frac{\partial}{\partial E}\left(\arg\left(A\right)\right) = \hbar\frac{\partial\alpha}{\partial E}$. Thus, the group delay in transmission is the same in either direction, or $\overline{\tau_{gt}} = \tau_{gt}$. Note that α here is equivalent to $\phi_T + kL$ in equation (4.20).

According to (4.28), the transmission probability $\mathbb{T}$ for the direct process would be the transmitted particle flux $\tilde{k}|F|^2$ divided by the incident particle flux $k|U|^2$, which evaluates to $\mathbb{T} = \frac{\tilde{k}}{k}\left|A\right|^2$. By the same argument, the transmission probability $\overline{\mathbb{T}}$ for the indirect process is $\overline{\mathbb{T}} = (k/\tilde{k})|\overline{A}|^2$. Equation (4.33) dictates that these two

transmission probabilities are equal,

$$\overline{\mathbb{T}} = \frac{k}{\tilde{k}} \left|\overline{A}\right|^2 = \frac{\tilde{k}}{k} \left|A\right|^2 = \mathbb{T}. \tag{4.35}$$

A similar relation can be obtained for the reflected beams, since

$$\overline{B} = -\frac{M_{12}}{M_{11}}, \quad B = \frac{M_{21}}{M_{11}} \tag{4.36}$$

and from (4.26) the reflection probabilities are related by

$$\begin{aligned} \overline{\mathbb{R}} = \left|\overline{B}\right|^2 &= \frac{M_{12}M_{12}^*}{M_{11}M_{11}^*} \\ &= \frac{M_{21}^*M_{21}}{M_{11}M_{11}^*} \\ &= \left|B\right|^2 = \mathbb{R}. \end{aligned} \tag{4.37}$$

It's clear from this that $|\overline{B}| = |B|$, and that $\overline{\beta} = \arg(\overline{B})$ and $\beta = \arg(B)$ do not share the simple relationship found for α and $\overline{\alpha}$ in (4.34). To find this relation, we can relate B and $\overline{B}$ directly by eqs. (4.30a) and (4.36).

$$B^* = \frac{M_{21}^*}{M_{11}^*} = \frac{M_{12}}{M_{11}^*} = \frac{M_{12}M_{11}}{M_{11}M_{11}^*} = -\overline{B}\frac{A^*}{A} \tag{4.38}$$

$$\overline{B} = -B^*\frac{A}{A^*} \tag{4.39}$$

Writing $A = |A|e^{i\alpha}$, $B = |B|e^{i\beta}$, and $\overline{B} = |\overline{B}|e^{i\overline{\beta}}$, this gives us

$$\left|\overline{B}\right| e^{i\overline{\beta}} = |B| \frac{|A|}{|A|} e^{i(\pm\pi - \beta + 2\alpha)} \tag{4.40}$$

Setting the phases on either side equal to one another, we have

$$\overline{\beta} = \pm\pi - \beta + 2\alpha \tag{4.41}$$

The sign ambiguity can be resolved by considering the case of a symmetric barrier. In this case, $\overline{\beta} = \beta$ and the phase equality simplifies to

$$\beta = \pm\frac{\pi}{2} + \alpha \tag{4.42}$$

A here is simply R in (4.12), and B is Te^{ikL} with T given by (4.13). From these, we can work out α and β:

$$\alpha = \arg\left(\frac{1}{g}\right) = \arg\left(g^*\right) = -\tan^{-1}\left[\Delta \tanh \kappa L\right] \tag{4.43a}$$

$$\begin{aligned}\beta &= \arg\left(\frac{1}{g}\right) + \arg(-i) = \alpha + \arg\left(e^{-i\pi/2}\right) \\ &= -\frac{\pi}{2} + \alpha\end{aligned} \tag{4.43b}$$

from which it is clear that the negative sign should be chosen in (4.41). Thus, we finally have a complete relation describing the phases in reflection and transmission from an asymmetric barrier

$$\overline{\beta} = -\pi - \beta + 2\alpha \tag{4.44}$$

We've already seen that $\tau_{gt} = \hbar\frac{\partial\alpha}{\partial E}$, and similar definitions can be made for the group delays in reflection from either side of the barrier, τ_{gr} and $\overline{\tau}_{gr}$. Differentiating (4.44) with respect to E, we have

$$\tau_{gt} = \frac{\tau_{gr} + \overline{\tau}_{gr}}{2}. \tag{4.45}$$

Thus, the transmission group delay is the arithmetic mean of the two reflection delays, which may be different from one another in the case of an asymmetric barrier. For a symmetric barrier, $\tau_{gr} = \overline{\tau}_{gr} = \tau_{gt}$, indicating that a particle will experience the same delay regardless of whether the particle is ultimately reflected or transmitted.

It should also be noted at this point that for an asymmetric barrier, we could

define a "bi-directional" group delay based on a weighted average of τ_{gt} and τ_{gr},

$$\tilde{\tau}_g = |A|^2 \tau_{gt} + |B|^2 \tau_{gr}. \tag{4.46}$$

This expression simplifies to $\tilde{\tau}_g = \tau_{gt} = \tau_{gr}$ for a symmetric barrier, as expected.

4.2.3 Phase time expressions for the group delay

Returning to our symmetric barrier problem once again, we will now calculate the group delay explicitly. First, we express the transmitted wave as

$$\Psi_{\mathrm{III}}(L) = Te^{ikL} = \frac{1}{g} = \frac{g^*}{gg^*} = |A|\, e^{i\alpha}. \tag{4.47}$$

We can then express α as

$$\begin{aligned} \tan\alpha = \frac{\mathrm{Im}\Psi_{\mathrm{III}}}{\mathrm{Re}\Psi_{\mathrm{III}}} = \frac{\mathrm{Im}(g^*)}{\mathrm{Re}(g^*)} = -\Delta \tanh\kappa L \\ \alpha = \arg(T) + kL = -\tan^{-1}\left(\Delta \tanh\kappa L\right) \end{aligned} \tag{4.48}$$

Note that some other sources use a different form for α, such as that found in [95]. However, it can be shown that this form is equivalent to ours:

$$\begin{aligned} \Phi = \frac{\pi}{2} - \tan^{-1}\left(-\frac{\coth\kappa L}{\Delta}\right) \\ \tan\left(\Phi - \frac{\pi}{2}\right) = (\Delta \tanh\kappa L)^{-1} \\ \tan\left(\Phi - \frac{\pi}{2}\right) = -(\tan\Phi)^{-1} = (\Delta \tanh\kappa L)^{-1} \\ \Phi = -\tan^{-1}\left(\Delta \tanh\kappa L\right). \end{aligned}$$

It will be advantageous to evaluate several energy derivatives we will encounter in

the derivation for readability:

$$\frac{\partial \kappa}{\partial E} = \frac{\sqrt{2m}}{\hbar}\frac{\partial}{\partial E}\left(V_0 - E\right)^{1/2} = -\frac{m}{\hbar^2 k} \tag{4.49a}$$

$$\frac{\partial k}{\partial E} = \frac{\sqrt{2m}}{\hbar}\frac{\partial}{\partial E}E^{1/2} = \frac{m}{\hbar^2 k} \tag{4.49b}$$

$$\begin{aligned}
\frac{\partial \Delta}{\partial E} &= \frac{\partial}{\partial E}\frac{1}{2}\left(\frac{\kappa^2 - k^2}{\kappa k}\right) \\
&= \frac{1}{\kappa k}\left(\kappa\frac{\partial \kappa}{\partial E} - k\frac{\partial k}{\partial E}\right) + \frac{\kappa^2 - k^2}{-2\kappa^2 k^2}\left(\frac{\partial \kappa}{\partial E}k + \kappa\frac{\partial k}{\partial E}\right) \\
&= \frac{-2m}{\hbar^2 \kappa k} - \frac{m}{2\hbar^2 \kappa k}\left(\frac{\kappa}{k} - \frac{k}{\kappa}\right)^2 \\
&= \frac{-m}{2\hbar^2}\frac{1}{\kappa k}\left(\frac{\kappa}{k} + \frac{k}{\kappa}\right)^2 \\
&= \frac{-2m\Delta'^2}{\kappa k \hbar^2}
\end{aligned} \tag{4.49c}$$

In addition, the following substitution based on our definition of α in 4.48 will frequently show up in the derivation.

$$\begin{aligned}
\tan\alpha &= -\Delta \tanh \kappa L \\
\cos^2\alpha &= \frac{1}{1^2 + (-\Delta \tanh \kappa L)^2} \\
&= \frac{1}{1 + \Delta^2 \tanh^2 \kappa L},
\end{aligned} \tag{4.50}$$

Now, we can evaluate $\tau_g = \tau_{gt}$ fairly easily:

$$\begin{aligned}
\tau_g &= \hbar\frac{\partial\alpha}{\partial E} = \hbar\frac{\partial}{\partial E}\left(-\tan^{-1}\left[\Delta\tanh\kappa L\right]\right) \\
&= -\frac{\hbar}{1+\Delta^2\tanh^2\kappa L}\frac{\partial}{\partial E}\left(\Delta\tanh\kappa L\right) \\
&= -\hbar\cos^2\alpha\left(\frac{\partial\Delta}{\partial E}\tanh\kappa L + \Delta\,\mathrm{sech}^2\,\kappa L\frac{\partial(\kappa L)}{\partial E}\right) \\
&= \frac{mL}{\hbar k}\frac{\cos^2\alpha}{2}\left(4\Delta'^2\frac{\tanh\kappa L}{\kappa L} + 2\frac{k}{\kappa}\Delta\,\mathrm{sech}^2\,\kappa L\right) \\
&= \frac{mL}{\hbar k}\frac{\cos^2\alpha}{2}\left[\left(\frac{\kappa}{k}+\frac{k}{\kappa}\right)^2\frac{\tanh\kappa L}{\kappa L} + \left(1-\frac{k^2}{\kappa^2}\right)\mathrm{sech}^2\,\kappa L\right]
\end{aligned} \tag{4.51}$$

Equation (4.51) represents our final form for the group delay τ_g. This expression, arrived at through the phase time interpretation, accurately represents the behavior of the peak of the particle wavefunction in region III for an incident particle described by equation (4.16). We will now demonstrate how this expression leads to an apparent case of superluminal propagation, and show why causality is not violated in the tunneling process.

4.3 The Hartman effect and superluminality

Hartman investigated this effect in 1962 in the context of tunneling diodes, calculating the expected electron tunneling delay through metal-insulator-metal sandwiches [96]. He observed that for a thin barrier the transmitted wave function had the same "shape" as the incident wave function but the peak was shifted, indicating a delay slightly larger than the "equal time," or the time it would take the particle to propagate through a length of free space equal to the barrier thickness L.

As the barrier length increased, the filtering action of the barrier led to a slight blue shift of the transmitted particle's energy distribution. However, this effect is negligible as long as the particle has a very narrow energy distribution centered well below the barrier potential. More importantly, he noted that the predicted delay

for such a particle saturated with increasing barrier length, at some point becoming smaller than the equal time. This saturation of the group delay eventually earned the "Hartman effect" moniker [97]. The effect is demonstrated in Figure 4.5, which plots the group delay given by equation (4.51) and the equal time as a function of barrier width for $E = V_0/2$. For small κL, corresponding to a thin barrier, the group delay exceeds the equal time. However, the group delay saturates as κL increases, and eventually crosses the strictly linear equal time.

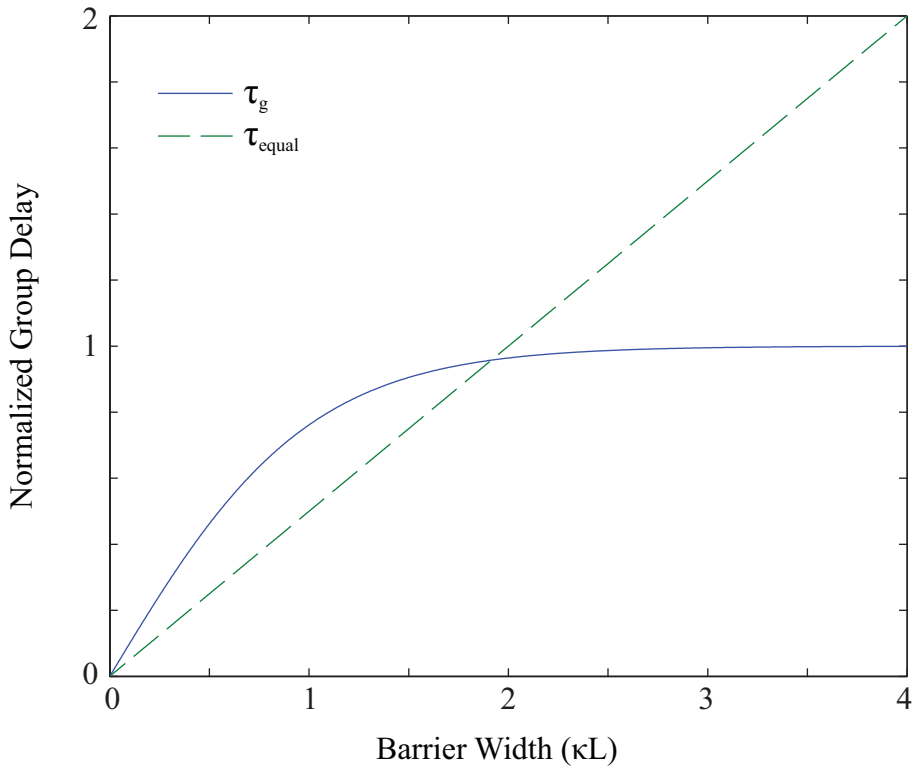

Figure 4.5: Group delay (solid line) and equal time L/v_0 (dashed line) as a function of barrier width κL. The equal time is the group delay for a particle traversing a distance L in the absence of a barrier. The delay normalization factor is $2/\kappa v_0$, and the particle energy is $E = V_0/2$. Adapted from [83].

When $E = V_0/2$, $k = \kappa$ and the second term in equation (4.51) vanishes, as does

α. In this special case, the expression for group delay simplifies to

$$\tau_g^{(E=V_0/2)} = \frac{2\tanh\kappa L}{\kappa v} \tag{4.52}$$

with $v = \hbar k/m$ as usual. It is very easy to see the mathematical source of the Hartman effect in this form. As $\kappa L \to \infty$, $\tanh\kappa L$ saturates to a value of 1 and the group delay saturates to $\tau_g^{E=V_0/2} \to 2/\kappa v$. The group delay should, therefore, become independent of length in the long-barrier limit. From our previous discussion, this also means that the reflection delay saturates to the same value for a symmetric barrier.

This limit is not even particularly difficult to achieve, as the saturation is almost complete at $\kappa L = 2$. Since $e^{-\kappa L}$ describes the decay of the wave function in the barrier region, one can intuitively interpret this condition as the point at which the wave function has decayed to $1/e^2$. While this interpretation is only an approximation, as it neglects the second term of Ψ_{II}, it corresponds to a barrier length of only a few particle wavelengths.

If we were to naïvely define a group velocity for the particle by dividing the barrier length by the group delay,

$$v_g = \frac{L}{\tau_g}, \tag{4.53}$$

we see the apparent conflict with causality. Once τ_g has saturated to a constant value, v_g becomes linear in L, which can grow without bound. At some value of L, v_g will exceed the speed of light which suggests superluminal propagation. The transmission of the particle decreases exponentially with the barrier length as well, leading to a further quandary: as the particle tunnels "faster," there's less and less of it to measure! In the infinite-barrier limit, the transmission drops to zero as v_g approaches infinity. These paradoxes have led to intense debate about the meaning of the group delay in barrier tunneling, and are in part responsible for the continued investigation of tunneling delays.

However, as we discussed in section 4.2, these paradoxes are based in the fun-

damental conceit that the group delay represents the propagation delay of a single unambiguous object, which is not the case. The peaks of the incident and transmitted particles are interference effects, and do not represent a localized physical object propagating from region I to region III.

It should also be noted that this is not an erroneous result caused by the use of a non-relativistic Schrödinger equation; the effect still occurs in the relativistic expression for group delay obtained from the Dirac equation [98]. It is a real effect that has been measured experimentally in both electromagnetic [99–106] and acoustic [107, 108] analogs to the barrier tunneling problem. According to Winful's fairly exhaustive summary of the experimental findings [83], the group delay exhibits all of the observable features discussed in this section. It accurately describes the time at which the transmitted peak exits the barrier, is shorter than the equal time for an appropriately long barrier, saturates with barrier length, and is equal in transmission and reflection for a symmetric barrier. The attenuated transmitted pulse is not reshaped or distorted as long as the tunneling occurs under quasi-static conditions. As such, any theoretical description of tunneling time needs to remain consistent with these findings.

We will now describe an alternative theory for tunneling delay that adheres to these conditions. This interpretation approaches the problem from a different perspective and provides additional physical intuition about the source and behavior of the delays predicted by the phase delay approach.

4.4 The dwell time interpretation of tunneling delay

We can also develop the group delay quantum-mechanically using a variational method outlined in [86, 109]. First, we note that we can define a "dwell time" τ_d as in [110].

$$\tau_d = \frac{\int_0^L |\Psi(x)|^2 \, dx}{j_{\text{in}}} \tag{4.54}$$

where $j_{\text{in}} = \hbar k/m$ is the incident particle flux. While this dwell time definition is an integral over a stationary state, we will now show that it is equivalent to the integral of the time-dependent wave function $\Psi(x,t)$ (normalized to unity),

$$I = \int_{-\infty}^{\infty} dt \int_0^L dx \, |\Psi(x,t)|^2 \, . \tag{4.55}$$

This integral in equation (4.55) can be arrived at by integrating the continuity equation over the barrier region as shown in [111, 112]. In the sources cited, the authors have assumed that the photon is far away from the barrier at time $t = 0$ so that $\Psi(x > 0, 0)$ is negligible, though this detail has no bearing on the following proof.

$\Psi(x,t)$ can be defined as a linear combination of stationary-state wave functions $\Psi_k(x)$,

$$\Psi(x,t) = \int \frac{dk}{2\pi} f(k) \, \Psi_k(x) \, e^{-iEt/\hbar}, \tag{4.56}$$

where $f(k)$ is the k-vector or frequency distribution of the input pulse, which we take to be the Fourier transform of the incident photon wavefunction $\Psi(x, 0)$:

$$f(k) = |f(k)| \, e^{i\xi(k)} = \int dx \, e^{-ikx} \, \Psi(x, 0). \tag{4.57}$$

We have assumed a transform-limited relationship here, though the relation is valid for pulses that are not transform-limited as long as $f(k)$ is unambiguously defined. The limits of integration in (4.56) have been omitted, but are assumed to cover the

regions of k-space where f is nonzero. $|\Psi(x,t)|^2$ is then

$$|\Psi(x,t)|^2 = \int \frac{dk}{2\pi} f(k)\,\Psi_k(x)\,e^{-iEt/\hbar} \int \frac{dk'}{2\pi} f(k')\,\Psi_{k'}(x)\,e^{-iE't/\hbar} \tag{4.58}$$

with E and E' being the energies that correspond to k and k' respectively. Then the integral I is

$$I = \int_{-\infty}^{\infty} dt \int_0^L dx \int\int \frac{dk}{2\pi}\frac{dk'}{2\pi} f(k)f^*(k')\,\Psi_k(x)\Psi^*_{k'}(x)\,e^{-i\hbar(k^2-k'^2)t/2m} \tag{4.59}$$

using the change of variables $q = k - k'$ and $Q = (k + k')/2$, such that $k = Q + q/2$ and $k' = Q - q/2$, this is

$$I = \int_{-\infty}^{\infty} dt \int_0^L dx \int\int \frac{dq}{2\pi}\frac{dQ}{2\pi} f(Q+q/2)f^*(Q-q/2) \Psi_{Q+q/2}(x)\Psi^*_{Q-q/2}(x)\,e^{-i\hbar qQt/m} \tag{4.60}$$

The time integral can be carried out to give

$$\int_{-\infty}^{\infty} dt\, e^{-i\hbar qQt/m} = \frac{2\pi m}{\hbar Q}\,\delta(q). \tag{4.61}$$

Substitution into I and evaluating the integral over q simplifies the expression further:

$$I = \int_0^L dx \int dq \int \frac{dQ}{2\pi}\left(\frac{m}{\hbar Q}\right)\delta(q)\,f(Q+q/2)f^*(Q-q/2) \Psi_{Q+q/2}(x)\Psi^*_{Q-q/2}(x) \tag{4.62}$$

$$= \int_0^L dx \int \frac{dQ}{2\pi}\left(\frac{m}{\hbar Q}\right) f(Q)f^*(Q)\Psi_Q(x)\Psi^*_Q(x) \tag{4.63}$$

From our earlier definitions of q and Q, we see that as $q \to 0$, $Q \to k$. Using the latter substitution, I becomes

$$I = \int \frac{dk}{2\pi}\,|f(k)|^2\,\frac{\int_0^L dx\,|\Psi_k(x)|^2}{\hbar k/m}. \tag{4.64}$$

Since $\Delta k << k$ for optical fields, the integral over dk acts like a delta function, leaving

$$I = \frac{\int_0^L dx\, |\Psi_k(x)|^2}{\hbar k/m}, \tag{4.65}$$

which is exactly the stationary-state integral definition of τ_d in equation (4.54).

Since we have expressions for $\Psi(x)$, C, and D within the barrier, we can evaluate τ_d directly from equation (4.54), noting that from equation (4.50)

$$\begin{aligned}
1/gg^* &= [(\cosh\kappa L + i\Delta \sinh\kappa L)(\cosh\kappa L - i\Delta\sinh\kappa L)]^{-1} \\
&= \left[\cosh^2\kappa L + \Delta^2 \sinh^2\kappa L\right]^{-1} \\
&= \left[\cosh^2\kappa L \left(1 + \Delta^2 \tanh^2\kappa L\right)\right]^{-1} \\
\frac{1}{gg^*} &= \frac{\cos^2\alpha}{\cosh^2\kappa L}.
\end{aligned}$$

with $\alpha = \phi_T + kL = -\tan^{-1}(\Delta \tanh\kappa L)$ as before).

$$\begin{aligned}
j_{\text{in}}\tau_d &= \int_0^L |\Psi(x)|^2\, dx \\
&= \frac{1}{2\kappa}\left[|C|^2\left(1 - e^{-2\kappa L}\right) + |D|^2\left(e^{2\kappa L} - 1\right)\right] + \left(C^*D + D^*C\right)L \\
&= \frac{1}{\kappa}\left[|C|^2 e^{-\kappa L} + |D|^2 e^{\kappa L}\right]\sinh\kappa L + \left(C^*D + D^*C\right)L \\
&= \frac{1}{4\kappa gg^*}\left[e^{\kappa L} + e^{-\kappa L}\right]\left(1 + \frac{k^2}{\kappa^2}\right)\sinh\kappa L + \frac{1}{2gg^*}\left(1 - \frac{k^2}{\kappa^2}\right)L \\
&= \frac{1}{2\kappa gg^*}\left[\left(1 + \frac{k^2}{\kappa^2}\right)\sinh\kappa L\cosh\kappa L + \kappa L\left(1 - \frac{k^2}{\kappa^2}\right)\right] \\
&= \frac{\cos^2\alpha}{2}\left[\left(1 + \frac{k^2}{\kappa^2}\right)\frac{\tanh\kappa L}{\kappa} + L\left(1 - \frac{k^2}{\kappa^2}\right)\operatorname{sech}^2\kappa L\right] \\
\tau_d &= \frac{mL}{\hbar k}\frac{\cos^2\alpha}{2}\left[\left(1 + \frac{k^2}{\kappa^2}\right)\frac{\tanh\kappa L}{\kappa L} + \left(1 - \frac{k^2}{\kappa^2}\right)\operatorname{sech}^2\kappa L\right]
\end{aligned} \tag{4.66}$$

It is clear that while strikingly similar, $\tau_d \neq \tau_g$ derived with the "phase time" interpretation. The relationship between the two can be calculated from the time-independent

Schrödinger equation for Ψ, $\hat{H}\Psi = E\Psi$. Differentiating, we get

$$\hat{H}\frac{\partial\Psi}{\partial E} = \Psi + E\frac{\partial\Psi}{\partial E} \tag{4.67}$$

note also that since E is real,

$$\hat{H}\Psi^* = E\Psi^* \tag{4.68}$$

We consider the equality

$$\Psi^* E\frac{\partial\Psi}{\partial E} - \frac{\partial\Psi}{\partial E}E\Psi^* = 0 \tag{4.69}$$

which is clearly true since Ψ, $\frac{\partial\Psi}{\partial E}$, and E are associative. Using (4.67) on the first term and (4.68) on the second, we have

$$\begin{aligned}\Psi^*\left(\hat{H}\frac{\partial\Psi}{\partial E} - \Psi\right) - \frac{\partial\Psi}{\partial E}\hat{H}\Psi^* &= 0\\ \Psi^*\Psi &= -\frac{\partial\Psi}{\partial E}\hat{H}\Psi^* + \Psi^*\hat{H}\frac{\partial\Psi}{\partial E}\end{aligned} \tag{4.70}$$

rearranging, and noting that $\hat{H} = \frac{-\hbar^2}{2m}\nabla^2$ and $\frac{\partial\Psi}{\partial E} = \frac{\partial\Psi}{\partial k}\frac{\partial k}{\partial E}$,

$$\begin{aligned}\Psi^*\Psi &= \frac{\hbar^2}{2m}\left[\frac{\partial\Psi}{\partial E}\frac{\partial^2\Psi}{\partial x^2} - \Psi^*\frac{\partial^3\Psi}{\partial^2 x\partial E}\right]\\ \Psi^*\Psi &= \frac{\hbar^2}{2m}\frac{\partial}{\partial x}\left[\frac{\partial\Psi}{\partial E}\frac{\partial\Psi^*}{\partial x} - \Psi^*\frac{\partial^2\Psi}{\partial x\partial E}\right]\\ \Psi^*\Psi &= \frac{\hbar^2}{2m}\frac{\partial}{\partial x}\left[\frac{\partial\Psi}{\partial k}\frac{\partial\Psi^*}{\partial x} - \Psi^*\frac{\partial^2\Psi}{\partial x\partial k}\right]\frac{\partial k}{\partial E}\end{aligned} \tag{4.71}$$

Integrating (4.71) over the barrier length gives us

$$\frac{2m}{\hbar^2}\int_0^L \Psi^*\Psi dx = \left[\frac{\partial\Psi}{\partial k}\frac{\partial\Psi^*}{\partial x} - \Psi^*\frac{\partial^2\Psi}{\partial x\partial k}\right]\Bigg|_0^L \frac{\partial k}{\partial E} \tag{4.72}$$

We can evaluate this expression at $x = 0$ and $x = L$ by using equations (4.4a) and

(4.4c). Since there is quite a bit of algebra, we'll do so in several steps.

$$\begin{aligned}\left(\frac{\partial\Psi}{\partial k}\frac{\partial\Psi^*}{\partial x}-\Psi^*\frac{\partial^2\Psi}{\partial k\partial x}\right)\Bigg|_{x=L} &= \\ &\left[\left(iLT+\frac{\partial T}{\partial k}\right)e^{ikL}\left(-ikT^*\right)e^{-ikL}-T^*e^{-ikL}\left(iT+ik\frac{\partial T}{\partial k}-kLT\right)e^{ikL}\right] \\ &=\left[kL|T|^2-ikT^*\frac{\partial T}{\partial k}-i|T|^2-ikT^*\frac{\partial T}{\partial k}+kL|T|^2\right] \\ &=-i2k\left[T^*\frac{\partial T}{\partial k}+\frac{|T|^2}{2k}+i|T|^2L\right] \end{aligned} \tag{4.73}$$

Expressing T as $|T|e^{i\phi_T}$ allows us to simplify further,

$$\begin{aligned}\left(\frac{\partial\Psi}{\partial k}\frac{\partial\Psi^*}{\partial x}-\Psi^*\frac{\partial^2\Psi}{\partial k\partial x}\right)\Bigg|_{x=L} &= \\ &-i2k\left[|T|e^{i\phi_T}\left(\frac{\partial|T|}{\partial k}e^{i\phi_T}+i|T|e^{i\phi_T}\frac{\partial\phi_T}{\partial k}\right)+\frac{|T|^2}{2k}+i|T|^2L\right] \\ &=-i2k\left[|T|\frac{\partial|T|}{\partial k}+\frac{|T|^2}{2k}+i|T|^2\left(\frac{\partial\phi_T}{\partial k}+L\right)\right] \\ &=-i2k\left[|T|\frac{\partial|T|}{\partial k}+\frac{|T|^2}{2k}+i\left(|T|^2\frac{\partial\alpha}{\partial k}\right)\right] \end{aligned} \tag{4.74}$$

where $\alpha=\phi_T+kL$ as before. Repeating this process for $x=0$ we have,

$$\begin{aligned}\left(\frac{\partial\Psi}{\partial k}\frac{\partial\Psi^*}{\partial x}-\Psi^*\frac{\partial^2\Psi}{\partial k\partial x}\right)\Bigg|_{x=0} &= \\ &\left[\frac{\partial R}{\partial k}(-ik)(1-R^*)-(1+R^*)\left(i(1-R)-ik\frac{\partial R}{\partial k}\right)\right] \\ &=\left[2ikR^*\frac{\partial R}{\partial k}-i\left(1-|R|^2\right)+i(R-R^*)\right] \\ &=i2k\left[R^*\frac{\partial R}{\partial k}-\frac{|T|^2}{2k}+i\frac{\mathrm{Im}(R)}{k}\right] \\ &=i2k\left[|R|e^{-i\phi_R}\left(\frac{\partial|R|}{\partial k}e^{i\phi_R}+i|R|e^{i\phi_R}\frac{\partial\phi_R}{\partial k}\right)-\frac{|T|^2}{2k}+i\frac{\mathrm{Im}(R)}{k}\right] \\ &=i2k\left[|R|\frac{\partial|R|}{\partial k}-\frac{|T|^2}{2k}+i\left(|R|^2\frac{\partial\phi_R}{\partial k}+\frac{\mathrm{Im}(R)}{k}\right)\right] \end{aligned} \tag{4.75}$$

Combining (4.72)-(4.75), we have,

$$\frac{2m}{\hbar^2}\int_0^L \Psi^*\Psi dx = -i2k\left[|T|\frac{\partial|T|}{\partial k}+|R|\frac{\partial|R|}{\partial k}\right]\frac{\partial k}{\partial E} + 2k\left[\left(|T|^2\frac{\partial\alpha}{\partial k}+|R|^2\frac{\partial\phi_R}{\partial k}+\frac{\mathrm{Im}(R)}{k}\right)\right]\frac{\partial k}{\partial E} \quad (4.76)$$

Here we note that since $|T|^2+|R|^2=1$, $\frac{\partial}{\partial k}(|T|^2+|R|^2)=0$, which eliminates the imaginary term. We can do some further simplification by moving terms around and using our previous definitions of j_{in}, τ_d, τ_{gt}, and τ_{gr}:

$$\begin{aligned}
\frac{m}{\hbar^2 k}\int_0^L \Psi^*\Psi dx - \frac{\mathrm{Im}(R)}{k}\frac{\partial k}{\partial E} &= \left(|T|^2\frac{\partial\alpha}{\partial k}+|R|^2\frac{\partial\phi_R}{\partial k}\right)\frac{\partial k}{\partial E}\\
\frac{1}{j_{\mathrm{in}}}\int_0^L \Psi^*\Psi dx - \frac{\mathrm{Im}(R)}{k}\hbar\frac{\partial k}{\partial E} &= \left(|T|^2\hbar\frac{\partial\alpha}{\partial E}+|R|^2\hbar\frac{\partial\phi_R}{\partial E}\right)\\
\tau_d - \frac{\mathrm{Im}(R)}{k}\hbar\frac{\partial k}{\partial E} &= \left(|T|^2\tau_{gt}+|R|^2\tau_{gr}\right)\\
\tau_d - \frac{\mathrm{Im}(R)}{k}\hbar\frac{\partial k}{\partial E} &= \tilde{\tau}_g
\end{aligned} \quad (4.77)$$

Thus, the dwell time τ_d and the bi-directional group delay $\tilde{\tau}_g$ differ by a term we define as the "self-interference delay" τ_i,

$$\tau_i = -\frac{\mathrm{Im}(R)}{k}\hbar\frac{\partial k}{\partial E} \quad (4.78)$$

The term "self-interference delay" seems to have been first used by Steinberg [91], though he did not define it mathematically. Note that τ_i can be, and often is, negative. It arises out of the overlap of incident and reflected waves in region I, in front of the barrier. This is reinforced by its disappearance when reflection coefficient becomes zero, such as at a barrier resonance or in the absence of a barrier. It can also be written as $\tau_i = -\hbar\,\mathrm{Im}(R)\frac{\partial}{\partial E}(\ln k)$, or noting that $\frac{\partial k}{\partial E} = m/\hbar^2 k$,

$$\tau_i = -\frac{\mathrm{Im}(R)}{kv} \quad (4.79)$$

where $v = j_{\text{in}} = \hbar k/m$. Winful notes in [86] that this has a very similar form to a one-dimensional scattering cross section $\sigma = 2\text{Im}(R)/k$, with scattering length $L_s = \text{Im}(R)/k$ divided by particle flux j_{in}. The self-interference delay can thus be interpreted as the time delay obtained from traversing that length. There is a notable similarity between this phenomenon and the optical theorem; both are caused by the interference of incident and scattered waves. Winful also notes that $\text{Im}(R)$ can be related to the Lagrangian for the Schrödinger equation in such a way as to identify it as a result of a stationary action principle, noting that for matter waves stationary action and stationary phase are equivalent concepts.

To provide yet another way to interpret the self-interference delay, Winful points out that it is common to define τ_d in classical contexts as $\tau_d = |T|^2\tau_T + |R|^2\tau_R$, treating transmission and reflection as mutually exclusive events. However, a quantum wave packet can be both transmitted and reflected simultaneously, and as such it is not proper to sum probabilities. If instead the summation is performed over complex amplitudes, the quantum-mechanical relation of dwell time to transmission and reflection delay works out to $\tau_d = |T|^2\tau_{gt} + |R|^2\tau_{gr} - \tau_i$, consistent with our current formulation. In this form, τ_i can be thought of as the delay caused by the interference of transmission and reflection amplitudes, which is neglected in the classical version.

Since this expression is directly obtained from the Schrödinger equation, it is more fundamental than the classical version. One could, of course, divide τ_i up in portions of $|T|^2$ and $|R|^2$ for a lossless barrier to put the latter into the form of the classical version, essentially defining $\tau_T = \tau_{gt} - \tau_i$ and $\tau_R = \tau_{gr} - \tau_i$. In that case, τ_T and τ_R become the Larmor times discussed in [110].

To evaluate this expression explicitly for the one-dimensional finite barrier prob-

lem, we use (4.12) to find Im(R):

$$\begin{aligned} R &= \left(\frac{1}{gg^*}\right)(-i\Delta' \sinh\kappa L)(\cosh\kappa L - i\Delta \sinh\kappa L) \\ &= \left(\frac{\cos^2\alpha}{\cosh^2\kappa L}\right)\left(-i\Delta' \sinh\kappa L \cosh\kappa L - \Delta\Delta' \sinh^2\kappa L\right) \\ &= -\left(\cos^2\alpha\right)\left(i\Delta' \tanh\kappa L + \Delta\Delta' \tanh^2\kappa L\right) \end{aligned}$$

$$\mathrm{Im}(R) = -\Delta' \cos^2\alpha \tanh\kappa L \tag{4.80}$$

Combining this expression with the explicit form of $\frac{\partial k}{\partial E}$ gives us

$$\begin{aligned} \tau_i &= -\frac{\hbar}{k}\left(-\Delta' \cos^2\alpha \tanh\kappa L\right)\left(\frac{m}{\hbar k}\right) \\ &= \frac{m}{\hbar k}\cos^2\alpha \frac{\Delta'}{k}\tanh\kappa L \\ &= \frac{m}{\hbar k}\frac{\cos^2\alpha}{2}\frac{1}{\kappa}\left(\frac{\kappa^2}{k^2}+1\right)\tanh\kappa L \\ &= \frac{mL}{\hbar k}\frac{\cos^2\alpha}{2}\left(1+\frac{\kappa^2}{k^2}\right)\frac{\tanh\kappa L}{\kappa L} \end{aligned} \tag{4.81}$$

From this, it is straightforward to show that $\tau_i + \tau_d$ as given in equations (4.66) and (4.81) is identical to the expression for τ_g in equation (4.51) by the phase time interpretation.

Figure 4.6 shows all three delay times as a function of normalized particle energy. At low energies ($E \ll V_0$), where most of the particle is reflected and $R \gg T$, the self-interference delay is the dominating component of the group delay. As particle energy increases, τ_i steadily decreases while the dwell time τ_d increases, eventually becoming the primary component at around $E = V_0/2$. At particle energies near $E = V_0$ or larger, where the reflected component is weak, the dwell time is the primary component of the group delay. When the particle experiences a transmission resonance, the self-interference delay becomes zero, consistent with a vanishing reflection coefficient.

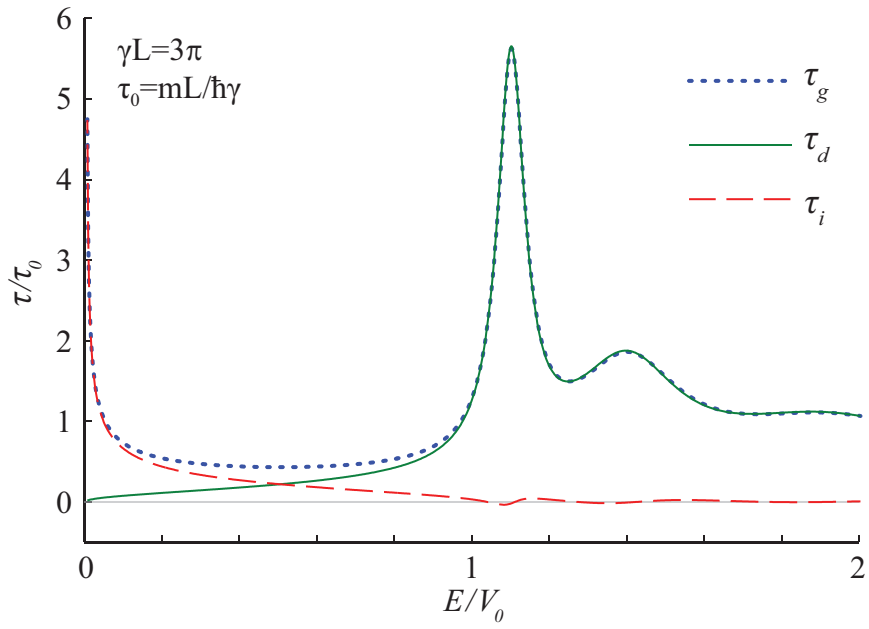

Figure 4.6: Group delay, dwell time, and self-interference delay as a function of normalized particle energy E/V_0 for a simple one-dimensional barrier. The delays are normalized to the "barrier time" $\tau_0 = L/v_0$, with $v_0 = \hbar\gamma/m$ and $\gamma = \sqrt{2mV_0}/\hbar$. This "barrier time" is the delay a particle of energy V_0 would experience when propagating through a distance L in free space. The barrier height in this plot is defined by $\gamma L = 3\pi$.

The group delay normalization shown in Figure 4.6 was chosen for consistency with [83,86]. However, this "barrier time" $\tau_0 = mL/\hbar\gamma$ may not be the most intuitive choice of normalization factor. It accurately describes the behavior of the group delay, which is large for $E \geq V_0$ and $E \to 0$ and small for $0 < E < V_0$. However, one could misinterpret this plot and conclude that the tunneling delay becomes slower than the vacuum propagation time when $E \to 0$, which is not the case.

This point is more clearly illustrated in Figure 4.7, which is normalized to the vacuum propagation time of the particle $\tau_{\text{vac}} = \hbar k/2m$. In this normalization scheme, we clearly see that even when $E \to 0$, the particle continues to tunnel through the distance L faster than it would propagate through that length. While the tunneling (group) delay does increase in this region just as we saw in Figure 4.6, the particle velocity in vacuum $v = \hbar k/m = \sqrt{2E/m}$ is also decreasing in this region, leading to

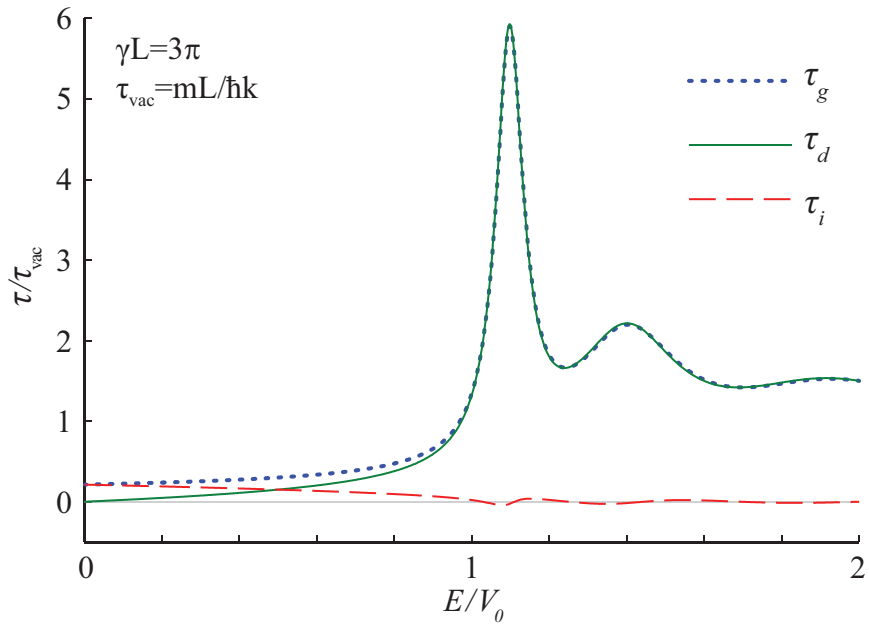

Figure 4.7: Group delay, dwell time, and self-interference delay as a function of normalized particle energy E/V_0 for a simple one-dimensional barrier. The delays are normalized to the "vacuum time" $\tau_0 = L/v$, with $v = \hbar k/m$, or the delay a particle of energy E would experience when propagating through a distance L in free space. The barrier height in this plot is defined by $\gamma L = 3\pi$, and $\gamma = \sqrt{2mV_0}/\hbar$ as before.

a longer vacuum propagation time. The vacuum propagation time continues to grow as fast or faster than the tunneling delay in this region, ensuring that the tunneling process continues to take less time than vacuum propagation.

As a final note, any constant normalization factor will generate a plot like Figure 4.6. This includes normalizing by a "causal time" $\tau_{\text{causal}} = L/c$, a case we are particularly interested in. On such a plot, the group delay would be superluminal in the vicinity of $E = V_0/2$ but become subluminal again as $E \to 0$. Yet we know that the delay should saturate as the barrier gets longer or higher ($\kappa L \to \infty$). These two seemingly disparate observations can be reconciled; the Hartman effect certainly still occurs at low particle energies, but the intersection of the group delay and causal time occurs at larger values of γL, the parameter which defines V_0 and L. If we assume that the barrier height V_0 (and thus γ) is fixed, this means that a longer barrier is

required to reach the superluminal-to-subluminal transition at smaller E.

The conceptual explanation is that the self-interference delay, which dominates in the low-E/V_0 region, is growing rapidly at low E/V_0 because the particle is spending more time "caught" in the standing wave at the barrier front. Since this growth occurs independently of L and V_0 as particle energy E decreases, the group delay is increasing while $\tau_{\text{causal}} = L/c$ is not, leading to a return to subluminal behavior. If we increase L without increasing V_0, E/V_0 is unchanged but τ_{causal} increases, which can push the system back into the superluminal regime.

4.5 Flux delays

The dwell time derived above is defined as the stored probability (or number of particles) within the barrier divided by the incident particle flux. It is essentially a measurement of the time it takes to build up the expected accumulation of stored particles in the barrier, or equivalently the time it takes to empty the barrier of stored particles when the incident flux stops. In each sense, it describes the escape process through the two exit channels of the barrier: transmission and reflection. The third escape channel, absorption, is ignored for the present discussion.

As outlined in §2.6 of [83], it is possible to consider the flux delay for each exit channel separately and relate those to τ_g. In this section, we will briefly reproduce this derivation for our one-dimensional barrier system.

The transmitted flux is canonically defined as

$$j_t = \text{Re}\left[\Psi^*(\hbar/im)\partial\Psi/\partial x\right], \tag{4.82}$$

with Ψ being the usual steady-state solution in a particular region. It can be verified that this is constant at any point x by evaluating it for Ψ_{I}-Ψ_{III} given by (4.4). In all three regions, j_t works out to $(\hbar k/m)|T|^2$. We could define a "transmitted flux

delay" by dividing the stored probability by the transmitted flux j_t,

$$\tau_{tf} = \frac{\int_0^L |\Psi(x)|^2}{j_t}. \tag{4.83}$$

It should be noted that this is not a transit time, which is perhaps easiest demonstrated by considering an example from fluid mechanics. Let a fluid have a local velocity field $v(x)$ such that the transmitted particle flux j_t depends on v and the local particle density $\rho(x) = \Psi^*(x)\Psi(x)$ as $j_t = \rho v$. We could define a delay time by considering the integral $\int dx/v$:

$$\int_0^L \frac{dx}{v(x)} = \frac{1}{j_t}\int_0^L \rho(x)dx = \frac{1}{j_t}\int_0^L \Psi^*(x)\Psi(x)dx \tag{4.84}$$

Note that this has the same form as (4.83). It is clear that this delay is made up of both forward- and backward-going components according to $v(x)$, and thus the delay is a property of the velocity field rather than any individual particle.

We can define an associated "reflected flux delay" in the same fashion,

$$\tau_{rf} = \frac{\int_0^L |\Psi(x)|^2}{|j_r|}, \tag{4.85}$$

with $j_r = -|R|^2\hbar k/m$ as calculated by substituting the reflected component of the wavefunction in region I, $\Psi_{\mathrm{I}}^{(R)}(x) = |R|e^{i\phi_r}e^{-ikx}$, into (4.82). Like τ_{tf}, τ_{rf} is not a transit time. Conceptually, either delay is more accurately thought of as the number of particles stored in the region $x = 0$ to $x = L$ divided by the rate at which particles leave that region via a given channel.

By inspection, the incident flux must be equal to the sum of transmitted and reflected fluxes, so $j_{\mathrm{in}} = j_t + |j_r|$. If we divide this equation by $\int_0^L |\Psi(x)|^2 dx$, we have

$$\frac{1}{\tau_d} = \frac{1}{\tau_{tf}} + \frac{1}{\tau_{rf}}. \tag{4.86}$$

This equation is the usual relation for escape rate addition, where the reciprocal of the total escape rate is equal to the sum of the reciprocals of the escape rates for each individual channel. To see how this relates to group delay, we combine the definition of bi-directional group delay in (4.46) with the expression $\tau_g = \tau_d + \tau_i$ and re-arrange to write

$$\tau_d = |T|^2 (\tau_{gt} - \tau_i) + |R|^2 (\tau_{gr} - \tau_i)$$

and since $\tau_d/|T|^2 = \tau_{tf}$ and $\tau_d/|R|^2 = \tau_{rf}$, we have

$$1 = \frac{\tau_{gt} - \tau_i}{\tau_{tf}} + \frac{\tau_{gr} - \tau_i}{\tau_{rf}}. \tag{4.87}$$

For a symmetric barrier, $\tau_{gt} = \tau_{gr} = \tau_g$ and $\tau_g - \tau_i = \tau_d$, and (4.87) reduces to (4.86).

4.6 Electromagnetic analog to tunneling

Up until now we have considered the one-dimensional barrier problem in the traditional context of particle wave functions. However, the concept of tunneling is not limited to particles. It occurs for waves of all types, including electromagnetic waves [113]. The derivation of the Hartman effect for electromagnetic waves is almost identical to that of particles because of the similarity between the time-independent Schrödinger equation and the Helmholtz equation. If we write the two side by side, the equivalence becomes clear.

$$\nabla^2 \tilde{\mathbf{E}} + \left(\frac{n(\bar{r})\omega}{c} \right) \tilde{\mathbf{E}} = 0 \tag{4.88a}$$

$$\nabla^2 \Psi + \frac{2m}{\hbar^2} (E - V(\bar{r})) \Psi = 0 \tag{4.88b}$$

If $k^2 = (n\omega/c)^2$ in equation (4.88a) is identical to $2m(E - V)/\hbar^2$ in equation (4.88b), then the electromagnetic wave $\tilde{E}$ will behave exactly the same way the quan-

tum mechanical wave function Ψ does in the finite barrier problem. In "forbidden" regions where k becomes imaginary, the electromagnetic wave will become evanescent and decay just like the particle wavefunction does within the barrier region. Practical situations where this occurs are common; undersized waveguides supporting no modes at frequency ω, photonic band gap structures, and frustrated total internal reflection are all examples of this behavior.

For the electromagnetic case, the dwell time is defined by average stored energy and input power rather than number of particles,

$$\tau_d = \frac{\langle U \rangle}{P_{\text{in}}} \tag{4.89}$$

where U and P_{in} are both averaged over a full cycle of the electromagnetic wave. The stored energy includes both electric and magnetic contributions:

$$\langle U \rangle = \langle U_{\text{e}} \rangle + \langle U_{\text{m}} \rangle$$

$$\langle U_{\text{e}} \rangle = \frac{1}{4} \int_V \mathbf{E} \cdot \mathbf{E}^* \frac{\partial}{\partial \omega} \left(\omega \epsilon(\omega) \right) dv \tag{4.90a}$$

$$\langle U_{\text{m}} \rangle = \frac{1}{4} \int_V \mathbf{H} \cdot \mathbf{H}^* \frac{\partial}{\partial \omega} \left(\omega \mu(\omega) \right) dv \tag{4.90b}$$

which need not be equal for an arbitrary structure. Usually the material dispersion is negligible compared to the waveguide's structural dispersion, so (4.90a)-(4.90b) simplify to

$$\langle U_{\text{e}} \rangle = \frac{1}{4} \int_V \epsilon \mathbf{E} \cdot \mathbf{E}^* dv \tag{4.91a}$$

$$\langle U_{\text{m}} \rangle = \frac{1}{4} \int_V \mu \mathbf{H} \cdot \mathbf{H}^* dv. \tag{4.91b}$$

Again, these stored energy terms include both transmitted and reflected components, and represent the wave as a whole rather than any individual result. As such, the dwell time calculated from these stored energy contributions is not a transit time and

cannot be associated with a velocity. The group delay in the electromagnetic case is defined exactly as in (4.20), or $\tau_g = \partial\phi/\partial\omega$.

To further illustrate this concept, Figure 4.8 shows the time-averaged energy density $\langle U \rangle$ as a function of position for two situations. In the top panel, we have a particle propagating through a length L of free space. The energy density in the region from $x = 0$ to $x = L$ (highlighted in green) is simply the usual free-space energy density $\langle U_0 \rangle$ of a propagating wave. In the second panel, we have replaced a section of free space with a barrier region that does not support propagating electromagnetic modes. The evanescent waves in this region penetrate a small distance into the barrier setting up an exponentially decaying energy density $\langle U \rangle < \langle U_0 \rangle$, highlighted in green. A significant amount of the wave is reflected, setting up a standing wave in the region $x < 0$, and a very small part of the wave is transmitted into region III (not visible).

We once again see how the Hartman effect manifests itself in this system. As the barrier length increases, the amount of stored energy in the free-space region increases roughly linearly, but the amount stored in the barrier configuration saturates very quickly. This reinforces the argument that the delay observed in tunneling is not a propagation delay but a cavity lifetime, describing the amount of time it takes for the cavity to respond to an increase in incident power.

Winful shows in [114] the case of the one-dimensional waveguide below cutoff, which is the most direct analog to the quantum-mechanical tunneling problem. Rather than present the full derivation here, we will briefly summarize the important points.

The system shown in Fig. 8(b) of [83], which is assumed to support a single TE_{10} mode, has electric fields $\mathbf{E} = \hat{y} \sin \eta x \Psi(z)$, with $\Psi(z)$ identical to that given in (4.4) for regions I-III other than a substitution of β for k. The variational theorem relates

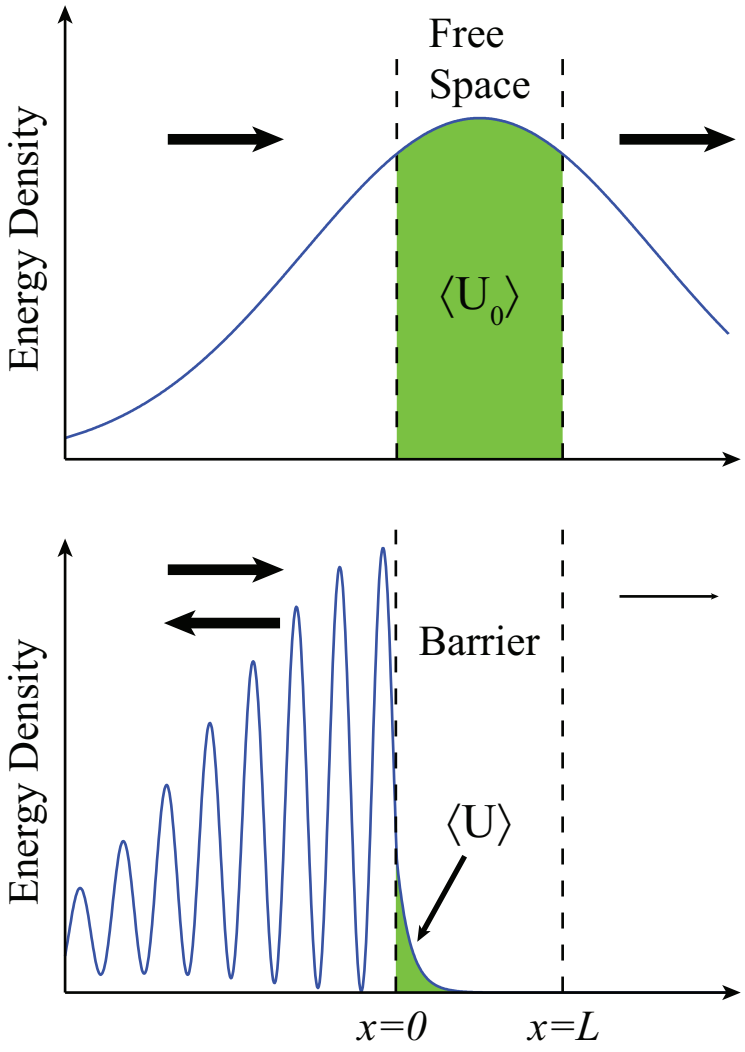

Figure 4.8: Plots of time-averaged energy density as a function of position for free-space and barrier regions. In the top panel, a pulse propagates through a length L of free space, and has the usual free-space energy density $\langle U_0 \rangle$. In the bottom panel, the section of free space is replaced with a barrier region, and the resulting energy density is significantly reduced by the exponential decay of the electric field in that region. The arrows in each panel represent the direction and relative magnitude of the plane-wave components $e^{\pm ikx}$ in each region.

E and **H** to the time average stored energy $\langle U \rangle$ in region II as follows:

$$\oint_S \left[\frac{\partial \mathbf{E}}{\partial \omega} \times \mathbf{H}^* + \mathbf{E}^* \times \frac{\partial \mathbf{H}}{\partial \omega} \right] d\mathbf{s} = 4i\langle U \rangle \tag{4.92}$$

The surface integral is carried out over the boundaries of region II. Since the waveguide walls are assumed to be metal, this simplifies the integral to an integration over the planes $z = 0$ and $z = L$. If we substitute our expression for **E** into the integrand, we have,

$$\oint_S \left(\frac{i|E_0|^2}{\omega \mu_0} \right) \left[\frac{\partial \Psi}{\partial \omega} \frac{\partial \Psi^*}{\partial z} - \Psi^* \frac{\partial^2 \Psi}{\partial \omega \partial z} + \frac{\Psi^*}{\omega} \frac{\partial \Psi}{\partial z} \right] \hat{\mathbf{z}} \cdot d\mathbf{s} = 4i\langle U \rangle \tag{4.93}$$

This looks strikingly similar to (4.71), but with an additional term caused by the dispersion relation for 3-dimensional electromagnetic waves. The inclusion of transverse dimensions in the model contributes an additional term to the dispersion, and thus to the self-interference delay as we will see shortly. If we substitute our values for **E**, Ψ_{I}, and Ψ_{III} into this equation to evaluate the integral and solve for $\tau_g = \partial \phi_0 / \partial \omega$, a lengthy series of algebraic operations brings us to the following expression for τ_g:

$$\tau_g \equiv \frac{d\phi_0}{d\omega} = \frac{\langle U \rangle}{P_{\mathrm{in}}} + \frac{\mathrm{Im}(R)}{\beta} \left(\frac{\beta}{\omega} - \frac{d\beta}{d\omega} \right). \tag{4.94}$$

Here $P_{\mathrm{in}} = \epsilon_0 |E_0|^2 A c^2 \beta / 4\omega$ is the time-averaged incident power, E_0 the amplitude of the incident mode in region I, A is the cross-sectional area of the waveguide in region I, and $\mathrm{Im}(R)$ is the imaginary part of the reflected wave amplitude in (4.4a). The first term is clearly the dwell time τ_d given in (4.89). The second term is the self-interference delay,

$$\tau_i = \frac{\mathrm{Im}(R)}{\beta} \left(\frac{\beta}{\omega} - \frac{d\beta}{d\omega} \right) \tag{4.95}$$

and is reminiscent of the same term in the quantum calculation. We have the same $\mathrm{Im}(R)$ factor that we saw in the quantum calculation, but also an extra dispersive factor that arises from the extra term in the variational theorem integrand. It's interesting to note that this dispersive factor arises solely from the evaluation of the

integral at $z = 0$, reinforcing its interpretation as a result of interference between incident and reflected waves in front of the barrier region. As before, the self-interference delay becomes zero if the reflection coefficient is zero or purely real. In addition, we see that it vanishes for a dispersionless waveguide, which has $d\beta/d\omega = \beta/\omega$. This is expected, as the interference pattern and phase fronts propagate at the same speed in such a waveguide, producing no extra delay. By applying the complex Poynting theorem [115], which in this case gives $\mathrm{Im}(R) = -\omega(\langle U_\mathrm{m}\rangle - \langle U_\mathrm{e}\rangle)/P_\mathrm{in}$, the self-interference delay can be written as

$$\tau_i = \frac{\langle U_\mathrm{m}\rangle - \langle U_\mathrm{e}\rangle}{P_\mathrm{in}} \left(\frac{v_{\mathrm{p}1}^0}{v_{\mathrm{g}1}^0} - 1\right) \tag{4.96}$$

where $v_{\mathrm{p}1}^0 = \omega/\beta$ and $v_{\mathrm{g}1}^0 = d\omega/d\beta$ are the phase and group velocities in the region before the barrier. $\langle U_\mathrm{m}\rangle - \langle U_\mathrm{e}\rangle$ is easily recognized as the reactive energy stored in the barrier region.

Winful notes [114] that this expression for $\mathrm{Im}(R)$ looks very similar to the definition of Q for a resonant cavity, which is

$$Q = \frac{\omega \times (\text{time-averaged energy stored in cavity})}{(\text{energy loss per second in cavity})}. \tag{4.97}$$

However, the form of (4.96) relates the reactive stored energy to the energy per second *incident* on the barrier. In that sense, it describes an "external" Q. As Winful puts it [114], "the barrier region forms an evanescent mode resonator with a finite decay time."

Explicit evaluation of (4.89) and (4.96) is straightforward and performed in [114] for a waveguide system of constant area A composed of a region of lower index n_2 sandwiched between two regions of index $n1 > n2$. We quote the results here for

completeness,

$$\tau_d = \frac{L}{v_{\mathrm{g1}}^0}\frac{\cos^2\phi_0}{2}\left[\frac{\omega_{\mathrm{c1}}^2}{\omega^2}\left(1+\frac{\beta^2}{\kappa^2}\right)\frac{\tanh\kappa L}{\kappa L} - \frac{n_2^2}{n_1^2}\left(\frac{\beta^2}{\kappa^2}-1\right)\operatorname{sech}^2\kappa L\right] \tag{4.98a}$$

$$\tau_i = \frac{L}{v_{\mathrm{g1}}^0}\frac{\cos^2\phi_0}{2}\frac{\omega_{\mathrm{c1}}^2}{\omega^2}\left(1+\frac{\kappa^2}{\beta^2}\right)\frac{\tanh\kappa L}{\kappa L} \tag{4.98b}$$

$$\tau_g = \frac{L}{v_{\mathrm{g1}}^0}\frac{\cos^2\phi_0}{2}\left[\frac{\omega_{\mathrm{c1}}^2}{\omega^2}\left(\frac{\beta}{\kappa}+\frac{\kappa}{\beta}\right)^2\frac{\tanh\kappa L}{\kappa L} - \frac{n_2^2}{n_1^2}\left(\frac{\beta^2}{\kappa^2}-1\right)\operatorname{sech}^2\kappa L\right] \tag{4.98c}$$

where $\omega_{\mathrm{c1}} = (n_2/n_1)\omega_{\mathrm{c2}}$ and $\omega_{\mathrm{c2}} = \eta c/n_2$ are the cutoff frequencies of the waveguide in regions I and II, which form the boundaries of the stop band. Note that for the electromagnetic case we will use ϕ_0 in place of α to conform to the notation found in other references. When evaluated in the limit of $L \to \infty$, these simplify to

$$\tau_d = \frac{2}{\kappa v_{\mathrm{g1}}^0}\left(\frac{\omega_{\mathrm{c1}}}{\omega}\right)^2\frac{\kappa^2}{\kappa^2+\beta^2} \tag{4.99a}$$

$$\tau_i = \frac{2}{\kappa v_{\mathrm{g1}}^0}\left(\frac{\omega_{\mathrm{c1}}}{\omega}\right)^2\frac{\beta^2}{\kappa^2+\beta^2} \tag{4.99b}$$

$$\tau_g = \tau_d + \tau_i = \frac{2}{\kappa v_{\mathrm{g1}}^0}\left(\frac{\omega_{\mathrm{c1}}}{\omega}\right)^2 \tag{4.99c}$$

Since all three of these become independent of length, all three delays demonstrate the Hartman effect.

4.7 Summary

In this chapter, we have reviewed the theory of the one-dimensional Hartman effect and described the predicted delays in terms of both a Wigner "phase time" interpretation as well as the "dwell time" interpretation promoted by Winful. We have shown that the two interpretations give consistent results, and provided some intuition about the meaning of the dwell time and self-interference delay. We have also shown how these concepts relate to flux delays, and demonstrated that the Hartman

effect has an electromagnetic analog which may be more convenient for experimental measurements.

The most important revelation in this chapter, however, is that the interpretation of the group delay as a transit time is incorrect because the peaks of the incident and transmitted particle are not indistinct causally related objects. As such, one cannot conclude that this quantity represents a propagation velocity or that a tunneled particle *travels* superluminally. At most, we can state that the particle's peak exhibits a delay that is shorter than the equivalent light propagation time in vacuum.

If we instead view the tunneling phenomenon in the dwell-time interpretation, we see the tunneling process in a different light, as a cavity effect. The delays observed in transmission and reflection are related to cavity lifetimes in this interpretation, which eliminates any apparent conflict with causality. In the next chapter, we will extend our investigation to two-dimensions systems and show that the dwell-time interpretation becomes even more relevant in that analysis.

Chapter 5

Tunneling delays in two dimensions

Tunneling delays and the Hartman effect can be observed in two-dimensional systems as well. Traditionally, the case considered is that of frustrated total internal reflection (FTIR). The group delay for FTIR has been worked out by several authors for the simple case of a glass-air-glass interface [82, 116, 117]. However, none of these treatments have addressed the problem in the context of dwell time and self-interference delay. Steinberg worked out expressions relating the group delay in FTIR to the group delay in the one-dimensional finite barrier problem [82], but was only able to provide these expressions in certain limits, for $E \ll V_0$, $E \approx V_0$, and $E \gg V_0$. In this chapter, we will demonstrate that by breaking the group delay down into its dwell time and self-interference components one can derive a single expression for group delay that is valid for arbitrary E. In addition, this expression illustrates the physical meaning behind the limiting cases observed by Steinberg.

5.1 1-D Hartman review

To set up the two-dimensional calculation, we briefly revisit the one-dimensional problem. To represent a localized electron wave packet, we chose to integrate over a narrow band of stationary states in (4.16). We then made the argument that the

peak of this wave packet occurs according to the energy derivative of the phase of the integrand. This statement is an invocation of the stationary phase approximation; put briefly, that the integral only gains appreciable contributions in locations where the derivative of the phase of the integrand is near zero. Since the integral in this case represents the output wave function, this is equivalent to stating that the wave packet's maximum occurs where all of the individual frequency components are in-phase with one another and interfere constructively. This is generally only a good approximation when the phase varies rapidly with energy, which is valid in this case.

It is clear from (4.16) that the maximum of the incident wave, ignoring the interference with the reflected wave, occurs at $x = 0$ at a time t given by

$$\begin{aligned} \frac{\partial}{\partial E}\left(\arg\Psi_E - Et/\hbar\right)\big|_{x=0} &= \left.\frac{\partial\phi}{\partial E}\right|_{x=0} - t/\hbar = 0 \\ t &= \hbar\left.\frac{\partial\phi}{\partial E}\right|_{x=0} \end{aligned} \tag{5.1}$$

where the phase derivative is evaluated only for the incident portion of the wavefunction in region I at $x = 0$, $\Psi_{\mathrm{I}}^{(\mathrm{in})} = e^{ikx}$. The choice of incident wave function and initial barrier plane ensure that $\phi = 0$ at $x = 0$ and subsequently that the incident peak arrives at $t = 0$.

We can repeat this calculation for $x = L$ using $\phi = \arg\left(\Psi_{\mathrm{III}}(L)\right) - Et/\hbar$, and get

$$\begin{aligned} \tau_g &= \hbar\frac{\partial\phi_L}{\partial E} = \hbar\frac{\partial}{\partial E}\left(\arg(T) + kL\right) \\ \tau_g &= \hbar\frac{\partial}{\partial E}\left(\phi_T + kL\right) \end{aligned} \tag{5.2}$$

which we have shown previously in (4.20). In this chapter, we will use τ_g exclusively to represent the delay calculated in the one-dimensional quantum calculation for consistency. Other sources use a different notation; in particular Steinberg chooses to use τ_e in [82] to signify that it is an electron (i.e particle) delay rather than an electromagnetic analog delay.

We will develop the two-dimensional case in a similar fashion to [82]. As we will see, the addition of a second dimension is a straightforward process and the two-dimensional wave functions closely resemble those of the one-dimensional problem. However, the addition of a second dimension causes the stationary phase approximation to change from a simple derivative to a gradient operator, significantly complicating the calculation of delay times.

5.2 Frustrated total internal reflection

The system we will consider in our two-dimensional calculation is shown in Figure 5.1. A slab of material with index n_2 and width L is sandwiched between two regions of material of index n_1. We will label the regions I, II, and III from left to right as in the one-dimensional calculation, with the region of index n_2 acting as the barrier region. A photon is incident on the interface between regions I and II with angle θ to the surface normal at position $x = 0$, $y = 0$. The dotted lines in the figure represent the planes $y = 0$ and $y = \Delta y$.

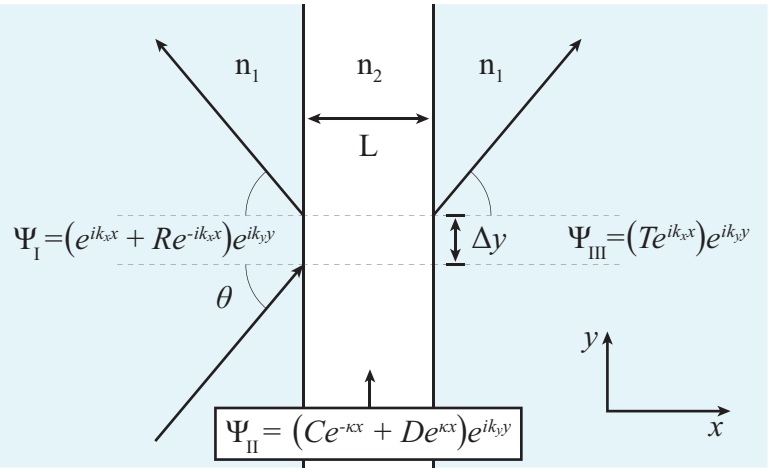

Figure 5.1: Schematic diagram for two-dimensional barrier tunneling.

For total internal reflection (TIR) to occur, Snell's law requires that $n_1 \sin\theta \geq n_2$, or that $\theta > \theta_c$, where $\theta_c = \arcsin(n_2/n_1)$ is the critical angle. In such a situation, the

majority of the wave will be reflected with a lateral Goos-Hänchen shift of Δy, while a small portion will tunnel to the other side and be transmitted, again with lateral shift Δy.

The electric field of a TE-polarized wave can be represented in regions I-III as

$$\mathbf{E}(x,y,t) = E(x,y,t)\hat{\mathbf{z}} = \Psi(x,y)e^{-i\omega t}\hat{\mathbf{z}} = \psi(x)e^{ik_y y - i\omega t} \tag{5.3}$$

where we have implicitly defined $k_i^2 = (n_i\omega/c)^2 = k_{ix}^2 + k_{iy}^2$ and taken advantage of the continuity of k_y across the barrier to eliminate the extra subscript. The system's symmetry dictates that $k_1 = k_3$ and $k_{1x} = k_{3x} = k_1\cos\theta$, which leaves us with only three unique components of k_i: k_{1x}, k_{2x}, and k_y. Substitution of this into the Helmholtz equation yields

$$\begin{aligned} \left(\nabla^2 + k_i^2\right)\psi e^{ik_y y} &= 0 \\ \psi'' + \left(k_i^2 - k_y^2\right)\psi &= 0 \end{aligned} \tag{5.4}$$

In regions I and III, this simplifies to

$$\psi'' + k_1^2\cos^2\theta\psi = 0,$$

while in region II the continuity of k_y means that we instead have

$$\psi'' + \left(k_2^2 - k_1^2\sin^2\theta\right)\psi = 0.$$

Or, substituting our definitions of k_i,

$$\text{I,III:} \qquad \psi'' + \frac{\omega^2}{c^2}n_1^2\cos^2\theta\psi = 0 \tag{5.5a}$$

$$\text{II:} \qquad \psi'' - \frac{\omega^2}{c^2}\left(n_1^2\sin^2\theta - n_2^2\right)\psi'' = 0 \tag{5.5b}$$

which bear a striking resemblance to (4.3) in the one-dimensional calculation. In fact, $\psi(x)$ has *exactly* the same solutions as the one-dimensional case. The Snell's law condition for TIR constrains $\left(n_1^2 \sin^2\theta - n_2^2\right)$ to be positive, forcing k_{2x} to be imaginary. To simplify the notation, we define $k_{2x} = i\kappa$ and drop the numerical subscript from k_{1x}, so that our three unique k_i components (k_{1x}, k_{2x}, k_y) become (k_x, $i\kappa$, k_y). $\Psi_{\mathrm{I}}(x,y)$ through $\Psi_{\mathrm{III}}(x,y)$ are then,

$$\Psi_{\mathrm{I}}(x,y) = \left(e^{ik_x x} + Re^{-ik_x x}\right) e^{ik_y y} \tag{5.6a}$$

$$\Psi_{\mathrm{II}}(x,y) = \left(Ce^{-\kappa x} + De^{\kappa x}\right) e^{ik_y y} \tag{5.6b}$$

$$\Psi_{\mathrm{III}}(x,y) = \left(Te^{ik_x x}\right) e^{ik_y y}. \tag{5.6c}$$

The terms in parentheses in (5.6) are $\psi(x)$ for regions I through III. If we plug these into (5.5), we find expressions for k_x and κ:

$$\text{I,III:} \qquad k_x^2 = \frac{\omega^2}{c^2} n_1^2 \cos^2\theta \tag{5.7a}$$

$$\text{II:} \qquad \kappa^2 = \frac{\omega^2}{c^2}\left(n_1^2 \sin^2\theta - n_2^2\right) \tag{5.7b}$$

Equation (5.7a) is simply our definition of k_x, while (5.7b) gives us the explicit definition of κ in terms of k and θ. We can compare this to the values of k and κ in the one-dimensional calculation in the text preceding equations (4.4a)-(4.4c), summarized in table 5.2 below.

Table 5.1: Coefficient Equivalences

	1-D (QM)	2-D (EM)
propagating wavevector k_x^2	$2mE/\hbar^2$	$\dfrac{n_1^2\omega^2}{c^2}\cos^2\theta$
evanescent wavevector κ^2	$2m(V_0 - E)/\hbar^2$	$\left(\dfrac{\omega}{c}\right)^2 (n_1^2 \sin^2\theta - n_2^2)$

The tunneling dynamics of the x-direction in the two-dimensional problem differ

only in the values of k_x and κ. At the critical angle θ_c, κ is zero, corresponding to the case of $E = V_0$ in the 1-D calculation. For angles greater than the critical angle we have tunneling, while angles less than θ_c lead to free propagation in the gap.

We have developed this formulation for a TE-polarized wave with $\mathbf{E} = E\hat{\mathbf{z}}$. The tangential component of the electric field E_z must be continuous across the boundary, forcing $\psi(x)$ to be continuous. Assuming $\mu = 1$ ensures that B_y is also continuous, leading to the continuity of $\psi'(x)$ as shown in section 5.4. Thus, in this special case, the boundary conditions are identical to those of the Schrödinger equation for the one-dimensional tunneling problem solved earlier.

However, the expressions in this section for $\Psi(x, y)$ are also valid for a TM-polarized wave, with $\mathbf{H} = H\hat{\mathbf{z}}$. In the TM case, the boundary conditions are slightly more complicated, necessitating a different definition of the coefficients R, C, D, and T. We will address these details in section 5.4, where we evaluate these coefficients for both polarizations.

Also note that the energy E and frequency $E/\hbar$ of the one-dimensional particle are *not* equal to the energy E_p and frequency ω of the photon in the two-dimensional case. We have assumed that the k-vector of the one-dimensional particle is equal to k_x in the two-dimensional case, and $E_p = \hbar\omega$ as usual.

5.3 Wave packets in two dimensions

To construct our wave packet in the two-dimensional case, we integrate over a narrow frequency band as before,

$$\begin{aligned} E(\bar{\mathbf{r}}, t) &= \int\int f(\bar{\mathbf{k}}) E_{\bar{\mathbf{k}}}(\bar{\mathbf{r}}, t) d\bar{\mathbf{k}} \\ &= \int\int f(\bar{\mathbf{k}}) \Psi_{\bar{\mathbf{k}}}(x, y) e^{-i\omega_{\bar{k}} t} d\bar{\mathbf{k}} \end{aligned} \tag{5.8}$$

where $f(\bar{\mathbf{k}})$ again represents a distribution of incident wavevectors, and $\Psi_{\bar{\mathbf{k}}}$ are the steady-state functions given in (5.6) for a particular input vector $\bar{\mathbf{k}} = (k_x, k_y)$. In the two-dimensional case, the stationary phase approximation requires that the gradient of the phase in k-space must vanish. Writing this out explicitly with $\Phi = \phi_T + k_x x + k_y y - \omega t$ representing the total phase of the steady-state contribution at $\bar{\mathbf{k}}$, we have

$$\frac{\partial \Phi}{\partial k_x}\hat{\mathbf{x}} + \frac{\partial \Phi}{\partial k_y}\hat{\mathbf{y}} = 0, \tag{5.9}$$

and since k_x and k_y are independent this equation implies that each term must go to zero individually.

Steinberg states in [82] that these derivatives can instead be taken with respect to the magnitude and direction of the k-vector in the paraxial approximation. While he does not elaborate on this point, we can make a reasonable guess about his method. Since the gradient must have the same value in any coordinate system, it is valid to consider derivatives in wavevector components $k_{x'}$ and $k_{y'}$ in a rotated coordinate system (x', y') instead of k_x and k_y. We choose a coordinate system rotated from the default axes (x, y) by an angle θ_0. In this coordinate frame, the wavevector k is decomposed into $k_{x'} = k\cos(\theta - \theta_0)$ and $k_{y'} = k\sin(\theta - \theta_0)$. We assume that θ_0 is chosen such that $\theta - \theta_0$ is small, and thus $k_{x'} \approx k$ and $k_{y'} \approx k(\theta - \theta_0)$. From this it is easy to show that $\partial\Phi/\partial k_{x'} \approx \partial\Phi/\partial k$ and $\partial\Phi/\partial k_{y'} \approx \partial\Phi/\partial\theta$. Thus, the partial derivative with respect to $k_{x'}$ is equivalent to differentiation with respect to k or ω, and the partial derivative with respect to $k_{y'}$ is equivalent to differentiation with respect to θ.

However, no approximation is necessary to reach this conclusion. The stationary phase argument is simply that $\nabla\Phi = 0$, and could easily be carried out in polar coordinates (k, θ) instead of Cartesian coordinates (k_x, k_y). Doing so immediately gives us

$$\frac{\partial \Phi}{\partial k}\hat{\mathbf{k}} + \frac{1}{k}\frac{\partial \Phi}{\partial \theta}\hat{\theta} = 0. \tag{5.10}$$

Since each term must still go to zero independently, we arrive at the conclusion without making any approximation. Thus, we can express the evolution of the wave packet peak with the expressions

$$\left.\frac{\partial \Phi}{\partial \omega}\right|_{\theta} = 0 \tag{5.11a}$$

$$\left.\frac{\partial \Phi}{\partial \theta}\right|_{\omega} = 0 \tag{5.11b}$$

where we have chosen to represent the wavevector derivative as a frequency derivative for consistency with [82].

If we substitute $\Phi = \phi_T + k_x x + k_y y + \omega t$ into these equations at time $t = \tau_\gamma$, and note that $\phi_T + k_x x$ is equivalent to ϕ_0 from our one-dimensional calculation, we have

$$\tau_\gamma = \left.\frac{\partial \phi_0}{\partial \omega}\right|_{\theta} + \frac{n_1}{c}\Delta y \sin\theta \tag{5.12a}$$

$$0 = \left.\frac{\partial \phi_0}{\partial \theta}\right|_{\omega} + \frac{n_1 \omega}{c}\Delta y \cos\theta \tag{5.12b}$$

Equation (5.12a) clearly shows that there are two contributions to the delay. The first term, $\tau_0 = \partial\phi_0/\partial\omega|_\theta$, is the frequency derivative analog to (4.20) and represents the time delay due to the longitudinal shift in x. The second term is proportional to Δy, which is related to the angular dispersion of the transmitted phase $\partial\phi_0/\partial\theta$ through (5.12b). We could interpret this term as the time delay due to the lateral shift in y.

This can be understood a little more clearly by considering Figure 5.2 which is taken from [82]. The total group delay τ_γ represents the delay between the "arrival" of a phase front at ($x = 0$,$y = 0$) and the "departure" of a phase front at ($x = L$,$y = \Delta y$). Since we're considering plane waves, we could equivalently say that this phase front departed from ($x = L$,$y = 0$) and propagated through an amount of free space $\Delta y \sin\theta$, acquiring a time delay of $\frac{n_1}{c}\Delta y \sin\theta$. This leaves $\tau_0 = \tau_\gamma - \frac{n_1}{c}\Delta y \sin\theta$ as the delay acquired between "arrival" at ($x = 0$,$y = 0$) and "departure" from

$(x = L, y = 0)$. From this perspective, τ_0 represents the delay due to tunneling from $(0, 0)$ to $(L, 0)$ and $\frac{n_1}{c}\Delta y \sin\theta$ is just a propagation delay caused by the Goos-Hänchen shift Δy.

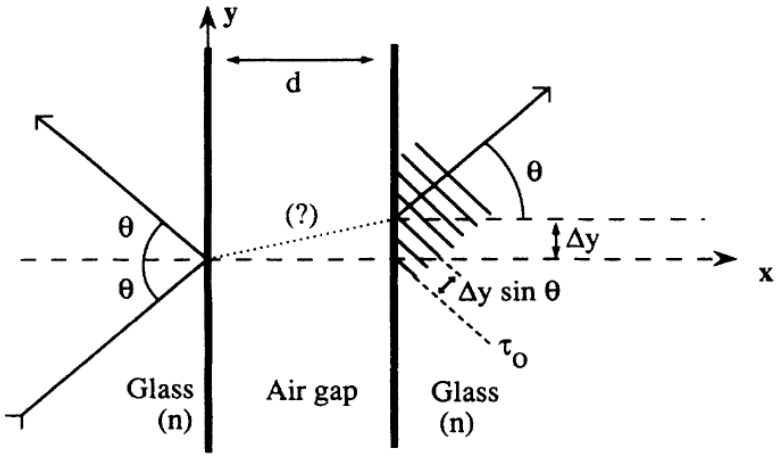

Figure 5.2: Diagram from [82] that illustrates the meaning of the second term in τ_γ.

The fact that τ_0 and Δy are independently determined by $\partial\phi_0/\partial\omega$ and $\partial\phi_0/\partial\theta$ respectively is interesting. In this case, it means that the x-direction dynamics are uniquely determined by the frequency derivative of ϕ_0 while the y-direction dynamics are uniquely determined by the angular derivative of ϕ_0. In addition, it suggests that if we can isolate τ_0 in a measurement, we would have a one-to-one correspondence with the one-dimensional problem.

We can combine equations (5.12) to express the group delay τ_γ for the electromagnetic case in terms of phase derivatives, which will allow us to relate τ_γ to the group delay τ_g of the tunneled particle in the one-dimensional calculation. For consistency, we will continue to use τ_d and τ_i to refer to the dwell time and self-interference delay of the one-dimensional calculation. In the same fashion as τ_γ, we will use Greek subscripts to indicate the electromagnetic versions of the dwell time and self-interference delay, τ_δ and τ_ι respectively.

Developing the relationships between τ_γ, τ_δ, τ_ι, τ_g, τ_d, and τ_i will be the focus of the next section.

5.4 Boundary conditions for TE and TM

The extension into two dimensions introduces an additional complication to the FTIR derivation of tunneling delay. The incident light now has two distinct polarization states to consider, and slight differences in the boundary conditions cause minor variations in the results. In most cases, these effects are very weak, and the TE derivation suffices to develop intuition. Most sources choose to deal with only the TE case [82, 116, 118, 119], though several address the differences in passing [120] or directly [117]. Our experiment had to be performed in the TM configuration, so we have developed a unified derivation that can properly represent either polarization state without difficulty. To our knowledge, this representation of the problem has not been presented before.

The bulk of the mathematical framework has already been set up in section 5.2. There, we developed expressions for an incident TE-polarized electric field $\mathbf{E} = E(x, y, t)\hat{\mathbf{z}}$ with a given input wavevector $\mathbf{k}$. These expressions have the form $E(x, y, t) = \psi(x)e^{ik_y y - i\omega t}$, where $\psi(x)$ is identical to the expressions found in the one-dimensional problem. However, as noted in that section, this treatment is also applicable to the case of a TM-polarized field if we instead consider $\mathbf{H} = H(x, y, t)\hat{\mathbf{z}}$, with $H(x, y, t) = \psi(x)e^{ik_y y - i\omega t}$.

To solve either the TE or TM case, one simply has to take the expressions for $\Psi(x, y)$ in equations (5.6) and use the appropriate boundary conditions to find explicit expressions for C, D, R, and T in terms of the fundamental quantities k, n_1, n_2, and L. We will now proceed with that derivation for each case, after which we will compare the two to demonstrate their similarity and present a notation which works properly for either case.

5.4.1 TE boundary conditions

For the TE case, the electric fields involved are

$$\mathbf{E}_{\mathrm{I}} = E_{\mathrm{I}}\hat{\mathbf{z}} = \left(e^{ik_x x} + Re^{-ik_x x}\right) e^{ik_y y - i\omega t}\hat{\mathbf{z}} \tag{5.13a}$$

$$\mathbf{E}_{\mathrm{II}} = E_{\mathrm{II}}\hat{\mathbf{z}} = \left(Ce^{-\kappa x} + De^{\kappa x}\right) e^{ik_y y - i\omega t}\hat{\mathbf{z}} \tag{5.13b}$$

$$\mathbf{E}_{\mathrm{III}} = E_{\mathrm{III}}\hat{\mathbf{z}} = \left(Te^{ik_x x}\right) e^{ik_y y - i\omega t}\hat{\mathbf{z}}. \tag{5.13c}$$

The boundary conditions are that the transverse components of $\mathbf{E}$ and $\mathbf{H}$, which for the TE case are E_z and H_y, are continuous across each interface. Continuity of E_z is straightforward, giving us two equations,

$$E_{\mathrm{I}}|_{x=0} = E_{\mathrm{II}}|_{x=0} \tag{5.14a}$$

$$E_{\mathrm{II}}|_{x=L} = E_{\mathrm{III}}|_{x=L} \tag{5.14b}$$

The continuity of H_y is identical to continuity of $\partial E_z/\partial x$ provided the magnetic permeability is invariant, or $\mu_{\mathrm{I}} = \mu_{\mathrm{II}} = \mu_{\mathrm{III}} = \mu$:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{5.15}$$

$$\begin{aligned} \frac{\partial E_z}{\partial y}\hat{\mathbf{x}} - \frac{\partial E_z}{\partial x}\hat{\mathbf{y}} &= i\omega\mathbf{B} \\ &= i\omega\mu\mathbf{H} \\ &= i\omega\mu\left(H_x\hat{\mathbf{x}} + H_y\hat{\mathbf{y}}\right) \end{aligned} \tag{5.16}$$

By inspection, the y-component must then satisfy

$$H_y = \frac{i}{\omega\mu}\frac{\partial E_z}{\partial x}. \tag{5.17}$$

Therefore, as long as μ is consistent across the boundaries, we can simplify our second boundary condition to the form

$$\left.\frac{\partial E_{\mathrm{I}}}{\partial x}\right|_{x=0} = \left.\frac{\partial E_{\mathrm{II}}}{\partial x}\right|_{x=0} \tag{5.18a}$$

$$\left.\frac{\partial E_{\mathrm{II}}}{\partial x}\right|_{x=L} = \left.\frac{\partial E_{\mathrm{III}}}{\partial x}\right|_{x=L}. \tag{5.18b}$$

These four boundary conditions are identical to the boundary conditions of Ψ given in (4.5) and (4.8) for the one-dimensional problem, and successfully demonstrate a one-to-one mapping of the two-dimensional problem on to the one-dimensional case. Unsurprisingly, if we evaluate these explicitly, we get equations (4.6) and (4.9),

$$1 + R = C + D \tag{5.19a}$$

$$\left(Ce^{-\kappa L} + De^{\kappa L}\right) = Te^{ik_x L}, \tag{5.19b}$$

$$ik\,(1 + R) = -\kappa\,(C - D) \tag{5.19c}$$

$$-\kappa\left(Ce^{-\kappa L} - De^{\kappa L}\right) = ik_x Te^{ik_x L}. \tag{5.19d}$$

From here, it is clear that the solutions for R, C, D, and T in terms of k, κ, and L will be identical to the one-dimensional case, with k and κ given by equations (5.7). To make it easier to spot the differences between TE and TM cases, we will repeat

the relevant equations here:

$$
\begin{aligned}
k_x^2 &= \frac{\omega^2}{c^2} n_1^2 \cos^2\theta \\
\kappa^2 &= \frac{\omega^2}{c^2}\left(n_1^2 \sin^2\theta - n_2^2\right) \\
\Delta &\equiv \frac{1}{2}\left(\frac{\kappa}{k} - \frac{k}{\kappa}\right) \\
\Delta' &\equiv \frac{1}{2}\left(\frac{\kappa}{k} + \frac{k}{\kappa}\right) \\
g &\equiv \cosh\kappa L + i\Delta \sinh\kappa L \\
R &= -\frac{i\Delta'}{g}\sinh\kappa L \\
T &= \frac{e^{-ikL}}{g} \\
C &= \left(1 - \frac{ik}{\kappa}\right) e^{\kappa L}/2g \\
D &= \left(1 + \frac{ik}{\kappa}\right) e^{-\kappa L}/2g
\end{aligned}
$$

5.4.2 TM boundary conditions

In the TM case, we instead focus on the magnetic field $\mathbf{H}$, which has only one component $H_z\hat{\mathbf{z}}$. The magnetic fields in each region can be expressed in similar fashion to (5.13),

$$\mathbf{H}_{\mathrm{I}} = H_{\mathrm{I}}\hat{\mathbf{z}} = \left(e^{ik_x x} + Re^{-ik_x x}\right) e^{ik_y y - i\omega t}\hat{\mathbf{z}} \tag{5.20a}$$

$$\mathbf{H}_{\mathrm{II}} = H_{\mathrm{II}}\hat{\mathbf{z}} = \left(Ce^{-\kappa x} + De^{\kappa x}\right) e^{ik_y y - i\omega t}\hat{\mathbf{z}} \tag{5.20b}$$

$$\mathbf{H}_{\mathrm{III}} = H_{\mathrm{III}}\hat{\mathbf{z}} = \left(Te^{ik_x x}\right) e^{ik_y y - i\omega t}\hat{\mathbf{z}}. \tag{5.20c}$$

The boundary conditions are still the continuity of all transverse components of **E** and **H**, but in the TM case this is the continuity of H_z and E_y. The first gives us

$$H_{\mathrm{I}}|_{x=0} = H_{\mathrm{II}}|_{x=0} \tag{5.21a}$$

$$H_{\mathrm{II}}|_{x=L} = H_{\mathrm{III}}|_{x=L}\,, \tag{5.21b}$$

exactly like E_z did in the TE case. Substitution of equations (5.20) into these expressions leads directly to equations (5.19a)-(5.19b).

Continuity of E_y is a little more subtle, however. Using the same logic as the TE case, we can start from $\nabla \times \mathbf{H} = \partial \mathbf{D}/\partial t$ to find that

$$E_y = -\frac{i}{\omega\epsilon}\frac{\partial H_z}{\partial x}. \tag{5.22}$$

However, it is clear that we cannot make the same assumption about the dielectric permittivity ϵ that we did about the magnetic permeability μ. The permittivity must change significantly at each interface to provide the necessary refractive index contrast to achieve FTIR. Thus, we require continuity of $(1/\epsilon)(\partial H_z/\partial x)$, or equivalently $(1/n^2)(\partial H_z/\partial x)$ for a lossless material, at each interface:

$$\left.\frac{1}{n_1^2}\frac{\partial H_{\mathrm{I}}}{\partial x}\right|_{x=0} = \left.\frac{1}{n_2^2}\frac{\partial H_{\mathrm{II}}}{\partial x}\right|_{x=0} \tag{5.23a}$$

$$\left.\frac{1}{n_2^2}\frac{\partial H_{\mathrm{II}}}{\partial x}\right|_{x=L} = \left.\frac{1}{n_1^2}\frac{\partial H_{\mathrm{III}}}{\partial x}\right|_{x=L}. \tag{5.23b}$$

While the conversion from ϵ to n^2 implicitly assumes that the materials in regions I-III are lossless, the results should retain validity for very weakly lossy materials. Plugging (5.20) into these equations gives the TM analogs to (5.19c) and (5.19d),

$$\frac{ik_x}{n_1^2}(1-R) = -\frac{\kappa}{n_2^2}(C-D) \tag{5.24a}$$

$$-\frac{\kappa}{n_2^2}\left(Ce^{-\kappa L} - De^{\kappa L}\right) = \frac{ik_x}{n_1^2}Te^{ik_x L} \tag{5.24b}$$

To simplify these expressions, we define a factor $M_{ij} = n_i^2/n_j^2$ which represents the refractive index contrast of the interface. Another way to interpret M_{ij} is the ratio of the intrinsic impedances of medium j to medium i, or $M_{ij} = (\eta_j^2/\eta_i^2)(\mu_i/\mu_j)$. Since our barrier is symmetric, we need only concern ourselves with $M_{12} = 1/M_{21}$. With this modification the four boundary conditions become,

$$1 + R = C + D \tag{5.25a}$$

$$\left(Ce^{-\kappa L} + De^{\kappa L}\right) = Te^{ik_x L} \tag{5.25b}$$

$$ik_x\,(1 - R) = -M_{12}\kappa\,(C - D) \tag{5.25c}$$

$$-M_{12}\kappa\left(Ce^{-\kappa L} - De^{\kappa L}\right) = ik_x Te^{ik_x L} \tag{5.25d}$$

As in the one-dimensional calculation, we combine equations (5.25a) and (5.25c) to express C and D in terms of R and T,

$$(5.25\text{a}) - \frac{1}{M_{12}\kappa}(5.25\text{c}) \Rightarrow \qquad 2C = 1 + R - \frac{ik_x}{M_{12}\kappa}\,(1 - R) \tag{5.26a}$$

$$(5.25\text{a}) + \frac{1}{M_{12}\kappa}(5.25\text{c}) \Rightarrow \qquad 2D = 1 + R + \frac{ik_x}{M_{12}\kappa}\,(1 - R) \tag{5.26b}$$

Using these expressions to eliminate C and D in (5.25b) gives

$$\begin{aligned} Te^{ik_x L} &= \frac{1}{2}\,(1 + R)\left(e^{\kappa L} + e^{-\kappa L}\right) \\ &\quad + \frac{ik_x}{2M_{12}\kappa}\,(1 - R)\left(e^{\kappa L} - e^{-\kappa L}\right) \\ Te^{ik_x L} &= (1 + R)\cosh\kappa L + \frac{ik_x}{M_{12}\kappa}\,(1 - R)\sinh\kappa L. \end{aligned} \tag{5.27}$$

And doing the same in (5.25d) yields

$$\begin{aligned} ik_x T e^{ik_x L} &= -\frac{M_{12}\kappa}{2}(1+R)\left(e^{-\kappa L} - e^{\kappa L}\right) \\ &\quad - \frac{ik_x}{2}(1-R)\left(-e^{-\kappa L} - e^{\kappa L}\right) \\ ik_x T e^{ik_x L} &= M_{12}\kappa(1+R)\sinh\kappa L + ik_x(1-R)\cosh\kappa L. \end{aligned} \tag{5.28}$$

We can now use equations (5.27) and (5.28) to eliminate T and solve for R,

$$\begin{aligned} (5.27) = \frac{1}{ik_x}(5.28) \Rightarrow& \\ (1+R)\cosh\kappa L + \frac{ik_x}{M_{12}\kappa}(1-R)\sinh\kappa L &= \frac{M_{12}\kappa}{ik_x}(1+R)\sinh\kappa L + (1-R)\cosh\kappa L \\ 2R\cosh\kappa L &= -i\left[\frac{k_x}{M_{12}\kappa}(1-R) + \frac{\kappa M_{12}}{k_x}(1+R)\right]\sinh\kappa L \\ 2R\cosh\kappa L &= -i\left[\left(\frac{k_x}{M_{12}\kappa} + \frac{M_{12}\kappa}{k_x}\right) + \left(\frac{\kappa M_{12}}{k_x} - \frac{k_x}{M_{12}\kappa}\right)R\right]\sinh\kappa L \\ 2R\cosh\kappa L &= -i\left[2\Delta_{\mathrm{M}}^{'} + 2\Delta_{\mathrm{M}}R\right]\sinh\kappa L \\ R(\cosh\kappa L + i\Delta_{\mathrm{M}}\sinh\kappa L) &= -i\Delta_{\mathrm{M}}^{'}\sinh\kappa L \\ R &= \frac{-i\Delta_{\mathrm{M}}^{'}\sinh\kappa L}{\cosh\kappa L + i\Delta_{\mathrm{M}}\sinh\kappa L} \\ R &= \frac{-i\Delta_{\mathrm{M}}^{'}}{g}\sinh\kappa L \end{aligned}$$

where we have used our earlier definition of $g = \cosh\kappa L + i\Delta\sinh\kappa L$, but replaced Δ and $\Delta^{'}$ with their TM-equivalents,

$$\begin{aligned} \Delta_{\mathrm{TM}} &\equiv \frac{1}{2}\left(\frac{M_{12}\kappa}{k_x} - \frac{k_x}{M_{12}\kappa}\right) \\ \Delta_{\mathrm{TM}}^{'} &\equiv \frac{1}{2}\left(\frac{M_{12}\kappa}{k_x} + \frac{k_x}{M_{12}\kappa}\right). \end{aligned}$$

Note that the expression for R is identical to both the TE and one-dimensional cases

outside of the substitution of Δ_{TM} for Δ and Δ'_{TM} for Δ'. In fact, with a simple redefinition of M_{ij},

$$M_{ij} \equiv \begin{cases} n_i^2/n_j^2 & \text{if TM-polarized,} \\ 1 & \text{otherwise (TE, 1-D).} \end{cases}$$

we find that $\Delta_{\mathrm{TM}} = \Delta$ and $\Delta'_{\mathrm{TM}} = \Delta'$ in the TE case. This definition of M_{ij} is equivalent to the factor m found in [120, 121] or the encapsulation of n_1 and n_2 into the particle effective mass in Lee's treatment [117]. With this updated definition of M_{ij}, we can dispense with the subscripts on Δ and have one consistent expression for R that applies to all three cases of interest.

With an explicit definition of R, we can now solve for T,

$$\begin{aligned}
&(5.27) + \frac{1}{ik_x}(5.28) \Rightarrow \\
2Te^{ik_xL} &= (1+R)\cosh\kappa L + \frac{ik_x}{M_{12}\kappa}(1-R)\sinh\kappa L \\
&\quad + \frac{M_{12}\kappa}{ik_x}(1+R)\sinh\kappa L + (1-R)\cosh\kappa L \\
&= 2\cosh\kappa L + i\left[\frac{k_x}{M_{12}\kappa}(1-R) - \frac{M_{12}\kappa}{k_x}(1+R)\right]\sinh\kappa L \\
&= 2\cosh\kappa L + i\left[\left(\frac{k_x}{M_{12}\kappa} - \frac{M_{12}\kappa}{k_x}\right) - R\left(\frac{k_x}{M_{12}\kappa} + \frac{M_{12}\kappa}{k_x}\right)\right]\sinh\kappa L \\
Te^{ik_xL} &= \cosh\kappa L - i\Delta\sinh\kappa L - iR\Delta'\sinh\kappa L \\
&= \cosh\kappa L - i\Delta\sinh\kappa L - i\left(\frac{-i\Delta'}{g}\right)\Delta'\sinh^2\kappa L \\
&= g^* - \frac{\Delta'^2}{g}\sinh^2\kappa L \\
&= \frac{1}{g}\left[gg^* - \Delta'^2\sinh^2\kappa L\right] \\
&= \frac{1}{g}\left[\cosh^2\kappa L + \Delta^2\sinh^2\kappa L - \Delta'^2\sinh^2\kappa L\right] \\
&= \frac{1}{g}\left[\cosh^2\kappa L - \sinh^2\kappa L\right] = \frac{1}{g}
\end{aligned}$$

$$T = \frac{e^{-ik_x L}}{g}.$$

Again, exactly like the previous cases. We can also find updated expressions for C and D, the process of which is made easier by recognizing a few simple relations,

$$\begin{aligned}
\Delta + \Delta^{'} &= \frac{M_{12}\kappa}{k_x} \\
\Delta - \Delta^{'} &= \frac{-k_x}{M_{12}\kappa} \\
1 \pm R &= g^{-1}\left[g \mp i\Delta^{'}\sinh\kappa L\right] \\
&= g^{-1}\left[\cosh\kappa L + i\left(\Delta \mp \Delta^{'}\right)\sinh\kappa L\right] \\
1 + R &= g^{-1}\left[\cosh\kappa L - i\frac{k_x}{M_{12}\kappa}\sinh\kappa L\right] \\
1 - R &= g^{-1}\left[\cosh\kappa L + i\frac{M_{12}\kappa}{k_x}\sinh\kappa L\right]
\end{aligned}$$

Plugging these expressions into (5.26) gives us expressions for C and D,

$$\begin{aligned}
2C &= 1 + R - i\frac{k_x}{M_{12}\kappa}(1 - R) \\
&= g^{-1}\left[\cosh\kappa L - i\frac{k_x}{M_{12}\kappa}\sinh\kappa L - i\frac{k_x}{M_{12}\kappa}\cosh\kappa L + \sinh\kappa L\right] \\
&= g^{-1}\left[\cosh\kappa L + \sinh\kappa L - i\frac{k_x}{M_{12}\kappa}(\cosh\kappa L + \sinh\kappa L)\right] \\
&= g^{-1}\left[e^{\kappa L} - i\frac{k_x}{M_{12}\kappa}e^{\kappa L}\right]
\end{aligned}$$

which simplifies to

$$C = \left(1 - i\frac{k_x}{M_{12}\kappa}\right)e^{\kappa L}/2g, \tag{5.30}$$

and a similar derivation yields

$$D = \left(1 + i\frac{k_x}{M_{12}\kappa}\right)e^{-\kappa L}/2g. \tag{5.31}$$

With this, our unified derivation is complete. C and D have a slightly different form than the earlier TE case because they now include the impedance factor M_{12}. Our updated definition of M_{12} accounts for this, so we can consider equations (5.30)-(5.31) the "proper" definitions of C and D, and the earlier expressions for the TE case a simplified special case.

5.4.3 Summary

We now have a general set of expressions for R, C, D, and T based on the boundary conditions of our two-dimensional tunneling problem, and can proceed with explicit calculation of the dwell time τ_δ, the self-interference delay τ_ι, and the group delay τ_γ. For convenience, we collect here all of the relevant equations we have derived that will be useful in the next section.

Wave function expressions

(time dependence $e^{-i\omega t}$ omitted)

1-D:	$\Psi_{\mathrm{I}} = e^{ikx} + Re^{-ikx}$	$\Psi_{\mathrm{II}} = Ce^{-\kappa x} + De^{\kappa x}$	$\Psi_{\mathrm{III}} = Te^{ikx}$
2-D TE:	$E_{\mathrm{I}} = \left(e^{ik_x x} + Re^{-ik_x x}\right) e^{ik_y y}$	$E_{\mathrm{II}} = \left(Ce^{-\kappa x} + De^{\kappa x}\right) e^{ik_y y}$	$E_{\mathrm{III}} = \left(Te^{ik_x x}\right) e^{ik_y y}$
2-D TM:	$H_{\mathrm{I}} = \left(e^{ik_x x} + Re^{-ik_x x}\right) e^{ik_y y}$	$H_{\mathrm{II}} = \left(Ce^{-\kappa x} + De^{\kappa x}\right) e^{ik_y y}$	$H_{\mathrm{III}} = \left(Te^{ik_x x}\right) e^{ik_y y}$

Coefficients

$$R = -i\frac{\Delta'}{g}\sinh\kappa L = |R|\, e^{i\phi_R} \qquad C = \left(1 - i\frac{k_x}{M_{12}\kappa}\right) e^{\kappa L}/2g$$

$$T = e^{ik_x L}/g = |T|\, e^{i\phi_T} \qquad D = \left(1 + i\frac{k_x}{M_{12}\kappa}\right) e^{-\kappa L}/2g$$

Definitions

$$
\begin{aligned}
g &\equiv \cosh \kappa L + i\Delta \sinh \kappa L \\
\Delta &\equiv \frac{1}{2}\left(\frac{M_{12}\kappa}{k_x} - \frac{k_x}{M_{12}\kappa}\right) \\
\Delta' &\equiv \frac{1}{2}\left(\frac{M_{12}\kappa}{k_x} + \frac{k_x}{M_{12}\kappa}\right) \\
M_{ij} &\equiv \begin{cases} n_i^2/n_j^2 & \text{if TM-polarized,} \\ 1 & \text{otherwise (TE, 1-D).} \end{cases} \\
\phi_0 &= \phi_T + k_x L
\end{aligned}
$$

5.5 Expressions for the dwell time, self-interference delay, and group delay

We proceed with the derivation in the TM case, as it is more general.

5.5.1 Dwell time

The general definition of dwell time from equation (4.54) is, in the notation of our two-dimensional problem,

$$j_{\text{in}}\tau_\delta = \int_0^L |\psi(x)|^2 \, dx,$$

where j_{in} remains the incident particle flux and $\psi(x)$ now replaces the one-dimensional wavefunction $\Psi(x)$. Since the field is not constrained in the y-dimension, the integration is still performed only over x. Another way to interpret this is that any effect on the dwell time due to the addition of the y-dimension is implicitly included in the calculation because as k_y varies, so does k_x and κ.

Substitution of $\psi(x)$ into this expression gives,

$$\begin{aligned} j_{\text{in}}\tau_\delta &= \int_0^L |C|^2 e^{-2\kappa x} + |D|^2 e^{2\kappa x} + CD^* + C^*D \\ &= -\frac{1}{2\kappa}\left(|C|^2 e^{-2\kappa x}\Big|_0^L + |D|^2 e^{2\kappa x}\Big|_0^L\right) + (CD^* + C^*D)\,L \\ &= \frac{1}{2\kappa}\left[|C|^2\left(1 - e^{-2\kappa L}\right) + |D|^2\left(e^{2\kappa L} - 1\right)\right] + (CD^* + C^*D)\,L \\ &= \frac{1}{\kappa}\left[|C|^2 e^{-\kappa L} + |D|^2 e^{\kappa L}\right]\sinh\kappa L + (CD^* + C^*D)\,L. \end{aligned}$$

Substituting our expressions for C and D gives us,

$$\begin{aligned} j_{\text{in}}\tau_\delta &= \frac{1}{4\kappa g g^*}\left[\left(1 + \frac{k_x^2}{M_{12}^2\kappa^2}\right)e^{\kappa L} + \left(1 + \frac{k_x^2}{M_{12}^2\kappa^2}\right)e^{-\kappa L}\right]\sinh\kappa L \\ &\quad + \frac{L}{4gg^*}\left[\left(1 - i\frac{k_x}{M_{12}\kappa}\right)^2 + \left(1 + i\frac{k_x}{M_{12}\kappa}\right)^2\right] \\ &= \frac{1}{4\kappa g g^*}\left[\left(e^{\kappa L} + e^{-\kappa L}\right)\left(1 + \frac{k_x^2}{M_{12}^2\kappa^2}\right)\right]\sinh\kappa L + \frac{L}{2gg^*}\left[1 - \frac{k_x^2}{M_{12}^2\kappa^2}\right] \\ &= \frac{1}{2\kappa g g^*}\left(1 + \frac{k_x^2}{M_{12}^2\kappa^2}\right)\sinh\kappa L\cosh\kappa L + \frac{L}{2gg^*}\left(1 - \frac{k_x^2}{M_{12}^2\kappa^2}\right) \end{aligned}$$

Remembering that

$$\frac{1}{gg^*} = \frac{\cos^2\phi_0}{\cosh^2\kappa L},$$

we can put the expression for the dwell time in a more recognizable form.

$$j_{\text{in}}\tau_\delta = L\frac{\cos^2\phi_0}{2}\left[\left(1 + \frac{k_x^2}{M_{12}^2\kappa^2}\right)\frac{\tanh\kappa L}{\kappa L} + \left(1 - \frac{k_x^2}{M_{12}^2\kappa^2}\right)\operatorname{sech}^2\kappa L\right] \tag{5.32}$$

Note that this is the same expression we have for τ_d in equation (4.66), but with additional factors of M_{12} to support two-dimensional TM boundary conditions. In fact, we can write τ_d in a slightly simpler form by using our definitions of Δ, Δ', and

$j_{\text{in}} = \hbar k/m$:

$$\tau_d = \frac{L\cos^2\phi_0}{j_{\text{in}}}\frac{k}{\kappa}\left[\Delta'\frac{\tanh\kappa L}{\kappa L} + \Delta\,\text{sech}^2\,\kappa L\right].$$

With a few subtle algebraic manipulations, we can put τ_δ in the same form,

$$\begin{aligned}\tau_\delta &= \frac{L\cos^2\phi_0}{2j_{\text{in}}}\frac{k_x}{M_{12}\kappa}\left[\left(\frac{M_{12}\kappa}{k_x}+\frac{k_x}{M_{12}\kappa}\right)\frac{\tanh\kappa L}{\kappa L}\right.\\ &\qquad\left.+\left(\frac{M_{12}\kappa}{k_x}-\frac{k_x}{M_{12}\kappa}\right)\text{sech}^2\,\kappa L\right]\\ \tau_\delta &= \frac{L\cos^2\phi_0}{j_{\text{in}}}\frac{k_x}{M_{12}\kappa}\left[\Delta'\frac{\tanh\kappa L}{\kappa L}+\Delta\,\text{sech}\,\kappa L\right]\end{aligned} \tag{5.33}$$

So the definition of dwell time carries over very well to the two-dimensional electromagnetic case. The only differences are an additional factor of $1/M_{12}$, similar variations within Δ and Δ', and of course a different definition of j_{in} than in the particle case. If we neglect the small variations in Δ and Δ', we can write,

$$\tau_\delta = \frac{j_{\text{in}}^{(1D)}}{M_{12}j_{\text{in}}}\tau_d \tag{5.34}$$

where we have used j_{in} to represent the incident flux in the general formulation, and $j_{\text{in}}^{(1\text{D})}$ to refer specifically to the incident particle flux of the one-dimensional electron-tunneling calculation.

5.5.2 Self-interference delay

To express τ_ι, we start with the definition in equation (4.79) with k_x substituted for k,

$$\tau_\iota = -\frac{\text{Im}(R)}{k_x j_{\text{in}}}.$$

We will later see that this definition is reasonable, as it gives us exactly the same result as the "phase time" development of τ_γ would suggest. For the moment, it will suffice to observe that the $\hbar\partial k/\partial E$ in (4.79) could be represented in the electromagnetic case as $\partial k/\partial\omega$, which evaluates to $1/v_{\text{group}}$. Since our materials are only weakly dispersive

for the photon bandwidths we are interested in, we can approximate this as n/c, or $1/v_{\text{phase}}$. The expression for $\text{Im}(R)$ is developed exactly as in equation (4.80),

$$\begin{aligned} R &= \left(\frac{g^*}{gg^*}\right)\left(-i\Delta^{'}\sinh\kappa L\right) \\ &= \left(\frac{\cos^2\phi_0}{\cosh^2\kappa L}\right)\left(-i\Delta^{'}\sinh\kappa L\cosh\kappa L - \Delta\Delta^{'}\sinh^2\kappa L\right) \\ &= -\left(\cos^2\phi_0\right)\left(i\Delta^{'}\tanh\kappa L + \Delta\Delta^{'}\tanh^2\kappa L\right), \end{aligned}$$

$$\text{Im}(R) = -\Delta^{'}\cos^2\phi_0\tanh\kappa L,$$

giving

$$\begin{aligned} \tau_\iota &= -\left(\frac{1}{k_x j_{\text{in}}}\right)\left(-\Delta^{'}\cos^2\phi_0\tanh^{\kappa L}\right) \\ &= \frac{L\cos^2\phi_0}{j_{\text{in}}}\frac{\kappa}{k_x}\left(\Delta^{'}\frac{\tanh\kappa L}{\kappa L}\right) \\ &= \frac{L}{j_{\text{in}}}\frac{\cos^2\phi_0}{2}\left(1+\frac{\kappa^2}{k_x^2}\right)\frac{\tanh\kappa L}{\kappa L} \end{aligned}$$

which is identical to equation (4.81). In this case, the only difference between τ_i and τ_ι is the choice of j_{in} such that

$$\tau_\iota = \frac{j_{\text{in}}^{(1\text{D})}}{j_{\text{in}}}\tau_i, \tag{5.35}$$

using the same definitions of j_{in} and $j_{\text{in}}^{(1\text{D})}$ as in (5.34).

5.5.3 Group delay

The phase of the transmitted field at $x = L$ is given by $H_{\text{III}}(x = L, y, t)$,

$$H_{\text{III}}(L, y, t) = Te^{ik_xL + ik_yy - i\omega t}. \tag{5.36}$$

The phase ϕ_H of this expression is simply

$$\phi_H = \phi_T + k_x L + k_y y - \omega t, \tag{5.37}$$

or $\phi_H = \phi_0 + k_y y - \omega t$ using our definition of ϕ_0. As we saw in section 5.3, the stationary phase condition gives us the two constraint equations,

$$\left(\frac{\partial \phi_H}{\partial \omega}\right)_\theta = 0$$
$$\left(\frac{\partial \phi_H}{\partial \theta}\right)_\omega = 0$$

which can be rewritten in the form of equations (5.12),

$$\tau_\gamma = \left(\frac{\partial \phi_0}{\partial \omega}\right)_\theta + \frac{n}{c}\Delta y \sin\theta \tag{5.39a}$$
$$0 = \left(\frac{\partial \phi_o}{\partial \theta}\right)_\omega + \frac{n\omega}{c}\Delta y \cos\theta \tag{5.39b}$$

If we eliminate Δy in these equations, we are left with

$$\tau_\gamma = \left(\frac{\partial \phi_0}{\partial \omega}\right)_\theta - \frac{\tan\theta}{\omega}\left(\frac{\partial \phi_0}{\partial \theta}\right)_\omega \tag{5.40}$$

We will now evaluate these expressions to represent τ_γ in terms of the dwell time τ_δ and self-interference delay τ_ι. It is slightly easier to do so if we expand the two derivatives of ϕ_0 in terms of k_x and κ derivatives. To make the algebra more transparent, we will first provide some useful identities.

$$k_x = \frac{n\omega}{c}\cos\theta \qquad\qquad \Delta \equiv \frac{1}{2}\left(\frac{M_{12}\kappa}{k_x} - \frac{k_x}{M_{12}\kappa}\right)$$

$$\kappa = \frac{\omega}{c}\sqrt{n_1^2\sin^2\theta - n_2^2} \qquad\qquad \Delta' \equiv \frac{1}{2}\left(\frac{M_{12}\kappa}{k_x} + \frac{k_x}{M_{12}\kappa}\right)$$

$$\phi_0 = -\arctan\Delta\tanh\kappa L \qquad\qquad M_{12} = \frac{n_1^2}{n_2^2}$$

$$\frac{\partial}{\partial x}(\arctan u) = \frac{1}{1+u^2}\frac{\partial u}{\partial x} \qquad\qquad \frac{1}{1+x^2} = \cos^2(\arctan x)$$

$$\frac{\partial\Delta}{\partial\kappa} = \frac{1}{2}\left(\frac{M_{12}}{k_x} + \frac{k_x}{M_{12}\kappa^2}\right) = \frac{1}{2\kappa}\left(\frac{M_{12}\kappa}{k_x} + \frac{k_x}{M_{12}\kappa}\right) = \frac{\Delta'}{\kappa}$$

$$\frac{\partial\Delta}{\partial k_x} = \frac{1}{2}\left(\frac{-M_{12}\kappa}{k_x^2} - \frac{1}{M_{12}\kappa^2}\right) = -\frac{1}{2k_x}\left(\frac{M_{12}\kappa}{k_x} + \frac{k_x}{M_{12}\kappa}\right) = -\frac{\Delta'}{k_x}$$

Since we'll be expanding in terms of κ and k_x derivatives, we will evaluate those explicitly as well,

$$\begin{aligned}
\left(\frac{\partial\phi_0}{\partial\kappa}\right)_{k_x} &= \frac{\partial}{\partial\kappa}\left(-\arctan\left(\Delta\tanh\kappa L\right)\right) \\
&= -\frac{1}{1+\Delta^2\tanh^2\kappa L}\frac{\partial}{\partial\kappa}\left(\Delta\tanh\kappa L\right) \\
&= -\cos^2\left(\arctan\left(\Delta\tanh\kappa L\right)\right)\left(\frac{\partial\Delta}{\partial\kappa}\tanh\kappa L + \Delta L\,\mathrm{sech}^2\,\kappa L\right) \\
&= -\cos^2\phi_0\left(\frac{\Delta'}{\kappa}\tanh\kappa L + \Delta L\,\mathrm{sech}^2\,\kappa L\right) \\
&= -L\cos^2\phi_0\left(\Delta'\frac{\tanh\kappa L}{\kappa L} + \Delta\,\mathrm{sech}^2\,\kappa L\right) \\
\left(\frac{\partial\phi_0}{\partial k_x}\right)_{\kappa} &= \frac{\partial}{\partial k_x}\left(-\arctan\left(\Delta\tanh\kappa L\right)\right) \\
&= -\cos^2\phi_0\frac{\partial}{\partial k_x}\left(\Delta\tanh\kappa L\right) \\
&= -\cos^2\phi_0\left(\frac{\partial\Delta}{\partial k_x}\tanh\kappa L + 0\right)
\end{aligned}$$

$$
\begin{aligned}
&= \cos^2\phi_0 \left(\frac{\Delta^{'}}{k_x}\right) \tanh\kappa L \\
\left(\frac{\partial k_x}{\partial \omega}\right)_\theta &= \frac{n_1}{c}\cos\theta = \frac{k_x}{\omega} \\
\left(\frac{\partial k_x}{\partial \theta}\right)_\omega &= -\frac{n_1\omega}{c}\sin\theta \\
\left(\frac{\partial \kappa}{\partial \omega}\right)_\theta &= \frac{1}{c}\sqrt{n_1^2\sin^2\theta - n_2^2} = \frac{\kappa}{\omega} \\
\left(\frac{\partial \kappa}{\partial \theta}\right)_\omega &= \frac{\omega}{2c}\left(n_1^2\sin^2\theta - n_2^2\right)^{-1/2}\left(2n_1^2\sin\theta\cos\theta\right) \\
&= \frac{\omega}{c}\left(\frac{\omega}{c\kappa}\right)\left(n_1^2\sin\theta\cos\theta\right) = \frac{n_1\omega}{c\kappa}\sin\theta\frac{n_1\omega}{c}\cos\theta \\
&= \frac{n_1\omega}{c}\frac{k_x}{\kappa}\sin\theta
\end{aligned}
$$

From here, we can evaluate the derivatives of ϕ_0 with respect to θ and ω found in equation (5.40).

$$
\begin{aligned}
\left(\frac{\partial \phi_0}{\partial \theta}\right)_\omega &= \left(\frac{\partial \phi_0}{\partial \kappa}\right)_{k_x}\left(\frac{\partial \kappa}{\partial \theta}\right)_\omega + \left(\frac{\partial \phi}{\partial k_x}\right)_\kappa \left(\frac{\partial k_x}{\partial \theta}\right)_\omega \\
&= \left[-L\cos^2\phi_0\left(\Delta^{'}\frac{\tanh\kappa L}{\kappa L} + \Delta\,\mathrm{sech}^2\,\kappa L\right)\right]\left(\frac{n_1\omega}{c}\frac{k_x}{\kappa}\sin\theta\right) \\
&\quad + \left(\cos^2\phi_0\frac{\Delta^{'}}{k_x}\tanh\kappa L\right)\left(-\frac{n_1\omega}{c}\sin\theta\right) \\
&= -L\frac{n_1\omega}{c}\sin\theta\cos^2\phi_0\left[\frac{k_x}{\kappa}\Delta^{'}\frac{\tanh\kappa L}{\kappa L} + \frac{k_x}{\kappa}\Delta\,\mathrm{sech}^2\,\kappa L + \frac{\Delta^{'}\kappa}{k_x}\frac{\tanh\kappa L}{\kappa L}\right] \\
&= -L\frac{n_1\omega}{c}\sin\theta\cos^2\phi_0\left[\frac{k_x}{\kappa}\left(\Delta^{'}\frac{\tanh\kappa L}{\kappa L} + \Delta\,\mathrm{sech}^2\,\kappa L\right) + \frac{\kappa}{k_x}\Delta\frac{\tanh\kappa L}{\kappa L}\right] \\
&= -L\frac{n_1\omega}{c}\sin\theta\left(M_{12}\tau_\delta + \tau_\iota\right)\frac{j_{\mathrm{in}}}{L} \\
&= -k_x j_{\mathrm{in}}\tan\theta\left(M_{12}\tau_\delta + \tau_\iota\right)
\end{aligned}
$$

$$\begin{aligned}
\left(\frac{\partial\phi_0}{\partial\omega}\right)_\theta &= \left(\frac{\partial\phi_0}{\partial\kappa}\right)_{k_x}\left(\frac{\partial\kappa}{\partial\omega}\right)_\theta + \left(\frac{\partial\phi_0}{\partial k_x}\right)_\kappa\left(\frac{\partial k_x}{\partial\omega}\right)_\theta \\
&= -L\cos^2\phi_0\left(\Delta'\frac{\tanh\kappa L}{\kappa L} + \operatorname{sech}^2\kappa L\right)\left(\frac{\kappa}{\omega}\right) + \cos^2\phi_0\frac{\Delta'}{k_x}\tanh\kappa L\left(\frac{k_x}{\omega}\right) \\
&= -\frac{\kappa^2 L}{k_x\omega}\left[\cos^2\phi_0\frac{k_x}{\kappa}\left(\Delta'\frac{\tanh\kappa L}{\kappa L} + \Delta\operatorname{sech}^2\kappa L\right)\right] \\
&\quad + \frac{Lk_x}{\omega}\left[\cos^2\phi_0\frac{\kappa}{k_x}\left(\Delta'\frac{\tanh\kappa L}{\kappa L}\right)\right] \\
&= -\frac{\kappa^2 L}{k_x\omega}\left[\frac{M_{12}j_{\text{in}}}{L}\tau_\delta\right] + \frac{Lk_x}{\omega}\left[\frac{j_{in}}{L}\tau_\iota\right] \\
&= -\frac{\kappa^2 j_{\text{in}}}{k_x\omega}M_{12}\tau_\delta + \frac{k_x j_{\text{in}}}{\omega}\tau_\iota
\end{aligned}$$

and plugging these into (5.40), we get our expression for τ_γ,

$$\begin{aligned}
\tau_\gamma &= \left(\frac{\partial\phi_0}{\partial\omega}\right)_\theta - \frac{\tan\theta}{\omega}\left(\frac{\partial\phi_0}{\partial\theta}\right)_\omega \\
&= \left[-\frac{\kappa^2 j_{\text{in}}}{k_x\omega}M_{12}\tau_\delta + \frac{k_x j_{\text{in}}}{\omega}\tau_\iota\right] - \frac{\tan\theta}{\omega}\left[-k_x j_{\text{in}}\tan\theta\left(M_{12}\tau_\delta + \tau_\iota\right)\right] \\
&= \frac{j_{\text{in}}}{k_x\omega}\left[\left(k_x^2\tan^2\theta - \kappa^2\right)M_{12}\tau_\delta + \left(k_x^2 + k_x^2\tan^2\theta\right)\tau_\iota\right] \\
&= \frac{j_{in}}{k_x}\frac{\omega}{c^2}\left[\left(n_1^2\sin^2\theta - n_1^2\sin^2\theta + n_2^2\right)M_{12}\tau_\delta + \left(n_1^2\cos^2\theta + n_1^2\sin^2\theta\right)\tau_\iota\right]
\end{aligned}$$

and finally,

$$\tau_\gamma = \frac{j_{\text{in}}\omega}{k_x c^2}\left[M_{12}n_2^2\tau_\delta + n_1^2\tau_\iota\right] \tag{5.41}$$

Let's scrutinize this result for a moment. It states that the observed group delay τ_γ can be decomposed into two components, one proportional to the dwell time τ_δ, and one proportional to the self-interference delay τ_ι. This seems reasonable and consistent with our one-dimensional results. In fact, we can relate τ_γ directly to τ_d and τ_i through equations (5.34) and (5.35). If we make those substitutions along with

the particle flux $j_{\text{in}}^{(1\text{D})} = \hbar k_x/m$, we find

$$\begin{aligned}\tau_\gamma &= \frac{j_{\text{in}}\omega}{k_x c^2}\left[M_{12}n_2^2\frac{j_{\text{in}}^{(1\text{D})}}{M_{12}j_{\text{in}}}\tau_d + n_1^2\frac{j_{\text{in}}^{(1\text{D})}}{j_{\text{in}}}\tau_i\right] \\ &= \frac{j_{\text{in}}^{(1\text{D})}\omega}{k_x c^2}\left[n_2^2\tau_d + n_1^2\tau_i\right] \\ &= \frac{\hbar k_x}{m}\frac{\omega}{k_x c^2}\left[n_2^2\tau_d + n_1^2\tau_i\right] \\ &= \frac{\hbar\omega}{mc^2}\left[n_2^2\tau_d + n_1^2\tau_i\right] \end{aligned} \tag{5.42}$$

This is a surprisingly simple mapping of the two-dimensional problem in k, θ, n_1, and n_2 onto the one-dimensional problem. In addition, it is applicable for arbitrary E and V_0 as implicitly defined by n_1, n_2, and $\bar{\mathbf{k}}$ in table 5.2,

$$V_0 = \frac{\hbar^2\omega^2}{2mc^2}\left(n_1^2 - n_2^2\right) \tag{5.43a}$$

$$E = V_0\frac{n_1^2\cos\theta}{n_1^2 - n_2^2} \tag{5.43b}$$

As a final confirmation, let us compare our results to those found in Steinberg and Chiao's analysis of this problem in [82]. In equation (21) of that paper, they define V_0 and E in terms of n, θ, and ω. The situation considered in their paper is a glass-air-glass system for which $n_1 = n$ and $n_2 = 1$. Under those conditions, our expressions (5.43) match theirs exactly. Note that their method of deriving τ_γ by breaking the phase derivatives in θ and ω down into derivatives in V_0 and E is entirely equivalent to our derivation, giving identical results.

Our general expression for τ_γ also matches theirs in the limits they consider. In the low-energy or "deep tunneling" limit ($k_x \ll \kappa$ or $E \ll V_0$), the bulk of the delay is due to the self-interference delay τ_i. Thus $\tau_d \to 0$ and $\tau_g \to \tau_i$, in which case (5.42) simplifies to

$$\tau_\gamma = \frac{\hbar\omega}{mc^2}n_1^2\tau_i, \tag{5.44}$$

in agreement with their result. Similarly in the "critical" limit, where $k \gg \kappa$ or $E \approx V_0$, $\tau_i \to 0$, and τ_d becomes the dominant contribution to group delay. This is also true of the "WKB" (Wentzel-Kramers-Brillouin) or "semiclassical" limit, where $E > V_0$ and κ becomes imaginary. Under either of those conditions, (5.42) simplifies to

$$\tau_\gamma = \frac{\hbar\omega}{mc^2} n_2^2 \tau_d, \tag{5.45}$$

again in complete agreement with their results since $n_2 = 1$.

They also observed that with a specific choice of effective mass for the photon, the one-dimensional tunneling delay seen by a massive particle will be identical to the tunneling delay seen by a photon. In the deep-tunneling limit (grazing incidence for the photon), the effective mass needs to be chosen such that $mc^2 = n_1^2\hbar\omega$, while in the critical and WKB limits that effective mass is reduced to $mc^2 = \hbar\omega$. Our expression clarifies this discrepancy by demonstrating that the appropriate expression in this limit is $mc^2 = n_2^2\hbar\omega$.

Our expression also provides intuition for why the effective mass changes. In the deep-tunneling limit, the particle delay is primarily due to self-interference before the barrier interface. In other words, the photon "spends more of its time" in a cavity-like state within a region of material with refractive index n_1. A similar argument shows that photons in the critical and WKB limits "spend more time" in the barrier region, which has a refractive index n_2.

Perhaps a more intuitive way to see this is to consider energies. In the cavity interpretation of tunneling, the delay is a representation of the amount of stored energy in the evanescent cavity. The energy density of an electromagnetic field is $u = \frac{1}{2}(\mathbf{E}\cdot\mathbf{D} + \mathbf{B}\cdot\mathbf{H})$, which will be proportional to $n^2\hbar\omega$ for a photon of frequency ω in a medium of index n. So we see that in any of the limits considered, the photon group delay τ_γ is simply the equivalent one-dimensional particle delay τ_g scaled by a ratio of energies - the energy of the photon and the rest mass of the particle mc^2. More generally, equation (5.42) could be interpreted to mean that the distribution of

delay between the self-interference and dwell-time is similar in the two cases, but each term is weighted individually by the permittivity of that region to properly adjust for the difference in energy density between a massive particle and electromagnetic waves.

It is interesting that both TE and TM cases map to the one-dimensional problem identically according to (5.42), despite having different boundary conditions. If we stick with the electromagnetic breakdowns τ_δ and τ_ι however, this is not the case. To demonstrate this, we substitute the electromagnetic particle flux $j_{\text{in}} = c/n$ into in equation (5.41) to find,

$$\begin{aligned}\tau_\gamma &= \frac{c}{n_1}\frac{1}{k_x}\frac{\omega}{c^2}\left[M_{12}n_2^2\tau_\delta + n_1^2\tau_\iota\right] \\ &= \frac{1}{n_1^2}\left[M_{12}n_2^2\tau_\delta + n_1^2\tau_\iota\right] \\ &= \left[M_{12}\frac{n_2^2}{n_1^2}\tau_\delta + \tau_\iota\right]. \end{aligned} \tag{5.46}$$

For the TM case, $\tau_\gamma = \tau_\delta + \tau_\iota$, which perfectly mimics the one-dimensional result $\tau_g = \tau_d + \tau_i$. The TE case differs only in the multiplicative factor n_2^2/n_1^2 in front of τ_δ. Surprisingly, in this form the TM case seems more similar to the one-dimensional result despite having different boundary conditions, as all of the dependence on n_1 and n_2 is encapsulated in the constituent delays τ_δ and τ_ι.

Another interesting observation we may make here is that a photon of arbitrary polarization undergoing FTIR will "break up," as its TE and TM components will experience different amounts of dwell time. Since $n_2 < n_1$ for the usual FTIR case, this suggests that the TE component will experience less total delay than the TM component.

5.6 Summary

In this chapter, we have recast the traditional two-dimensional FTIR problem in terms of a cavity-like model, with the total group delay divided into self-interference and dwell-time terms. We have shown that once expressed in these terms, the 2-D problem maps directly to the one-dimensional quantum-mechanical problem in a very simple fashion. In this form, the mapping is valid for arbitrary particle energy E and barrier height V_0, as defined in terms of n_1, n_2, θ, and k by similarity between the time-independent Schrödinger equation and Helmholtz equation. Further, we have shown that our version simplifies to match previous predictions in the special cases considered by Steinberg and Chiao in [82].

In addition, this interpretation of the tunneling delay phenomenon gives us new physical insight into the process. Conceptually, it breaks the process down into delays due to a cavity-like effect in the tunneling region and a standing-wave effect in the region of incidence. It also suggests that this is not an arbitrary decomposition, but rather that the dwell time and self-interference delays are physically meaningful quantities that may even be able to be tested and measured individually. In the next chapter, we will describe our attempts to measure the group delay in FTIR, and how those measurements necessarily differ from the ideal 2-dimensional case that we have presented here.

Chapter 6

Experimental measurement of tunneling delay in FTIR

In this chapter, we will discuss the double-prism (DP) system as a method of measuring tunneling delay. We will address the limitations of the device and the measurements we make with it, as well as the particular implementation we used to make time-delay measurements of barrier tunneling with a Hong-Ou-Mandel interferometer. Finally, we will present the results of those measurements and comment on the meaning of those results.

6.1 Double-prism systems and their limitations

The theoretical treatment presented in the previous chapter addressed the general case of tunneling delay in a two-dimensional FTIR structure. This approach is consistent with the usual treatment of the topic [82, 116, 117]. However, experimental measurement of this delay is slightly more complicated, as one has to consider the geometry of the experiment as a whole. The usual experimental implementation mimics that proposed in [82] and shown in Figure 6.1, which is a system of two prisms with an air gap. The theory presented in the literature does not accurately reflect the

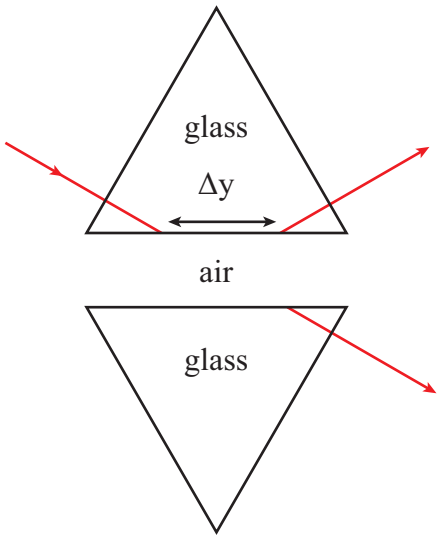

Figure 6.1: Experimental FTIR implementation as described in [82].

experimental measurements one can obtain in this system, because the practical issue of coupling in and out of the double-prism system is generally overlooked or ignored.

To illustrate this effect, we present Figure 6.2, which shows a simple diagram of FTIR in an equilateral prism. If the second region were filled with a perfectly conducting material, the light would reflect along the dotted red line, with no Goos-Hänchen shift Δy. However, if region II is air, we observe FTIR with a non-zero Goos-Hänchen shift and a simultaneous reduction in the amount of glass propagation length of the reflected beam. The amount of glass reduction σ can be related to Δy very simply,

$$\sigma = \Delta y \sin\theta, \tag{6.1}$$

and the total propagation delay for this amount of glass is $n\sigma/c = n\Delta y \sin\theta/c$. Thus, the delay we measure experimentally in an FTIR configuration is not the τ_γ derived in the previous chapter and in the bulk of the literature, but $\tau_{\gamma,\text{meas}} = \tau_\gamma - n\sigma/c$. By

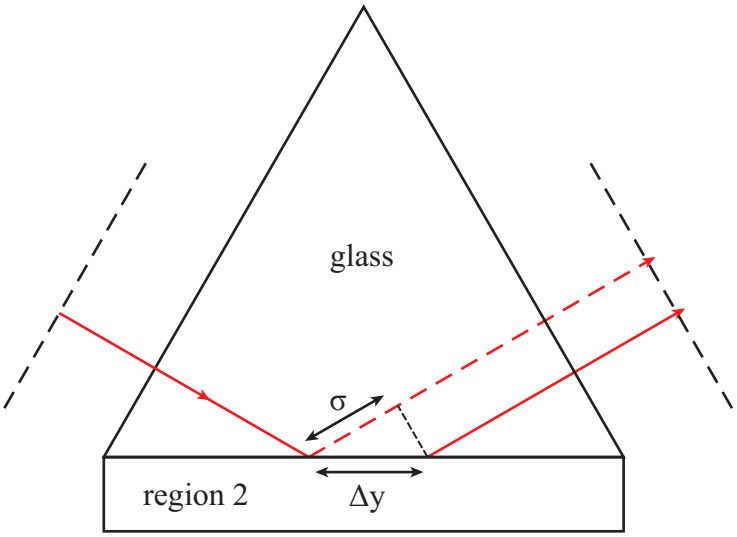

Figure 6.2: FTIR in an equilateral prism. The Goos-Hänchen shift Δy causes a change σ in the glass propagation length, which has ramifications for an experimental measurement of tunneling delay in the double-prism geometry.

inspection of equation (5.39), we see that

$$\begin{aligned}
\tau_{\gamma,\mathrm{meas}} &= \tau_\gamma - n\sigma/c \\
\tau_{\gamma,\mathrm{meas}} &= \left(\frac{\partial\phi_0}{\partial\omega}\right)_\theta + \frac{n}{c}\Delta y \sin\theta - \frac{n}{c}\Delta y \sin\theta \\
\tau_{\gamma,\mathrm{meas}} &= \left(\frac{\partial\phi_0}{\partial\omega}\right)_\theta .
\end{aligned} \tag{6.2}$$

As discussed in the previous chapter, the first term $\partial\phi_0/\partial\omega$ describes the delay contribution from the propagation in the x-direction while the second term describes the contribution from the y-direction, or Goos-Hänchen shift. This is perhaps more

clearly seen with the same substitution made immediately preceding equation (5.41):

$$\begin{aligned}\tau_{\gamma,\text{meas}} &= \left(\frac{\partial\phi_0}{\partial\omega}\right)_\theta \\ &= \left[-\frac{\kappa^2 j_{\text{in}}}{k_x\omega}M_{12}\tau_\delta + \frac{k_x j_{\text{in}}}{\omega}\tau_\iota\right] \\ &= \frac{j_{\text{in}}}{k_x\omega}\left[k_x^2\tau_\iota - \kappa^2 M_{12}\tau_\delta\right] \end{aligned} \tag{6.3}$$

Note that if we're below the critical angle, $\kappa = ik_{2x}$ and $-\kappa^2 = k_{2x}^2$, where we've used k_{2x} to represent the x-component of the propagation vector in the second material. One can express the "unmeasurable" portion in a similar fashion,

$$\begin{aligned}\tau_{\gamma,\text{unmeas}} &= -\frac{\tan\theta}{\omega}\left[-k_x j_{\text{in}}\tan\theta\,(\tau_\iota + M_{12}\tau_\delta)\right] \\ &= \frac{j_{\text{in}}}{k_x\omega}\left[k_x^2\tan\theta(\tau_\iota + M_{12}\tau_\delta)\right] \\ &= \frac{j_{\text{in}}}{k_x\omega}\left[k_y^2(\tau_\iota + M_{12}\tau_\delta)\right] \end{aligned} \tag{6.4}$$

We can also express the measurable and unmeasurable portions in terms of the 1-dimensional delays τ_d and τ_i using equations (5.34) and (5.35).

$$\tau_{\gamma,\text{meas}} = \frac{j_{\text{in}}^{(1\text{D})}}{k_x\omega}\left[k_x^2\tau_i - \kappa^2\tau_d\right] = \frac{\hbar}{m\omega}\left[k_x^2\tau_i - \kappa^2\tau_d\right] \tag{6.5a}$$

$$\tau_{\gamma,\text{unmeas}} = \frac{j_{\text{in}}^{(1\text{D})}}{k_x\omega}\left[k_y^2\,(\tau_i + \tau_d)\right] = \frac{\hbar}{m\omega}\left[k_y^2\,(\tau_i + \tau_d)\right] \tag{6.5b}$$

In other words, apart from an constant factor, the "measurable" portion of the delay is the weighted average of the dwell and self-interference delays, with each component weighted by the x-component of the wavevector in the appropriate region. The "unmeasurable" portion is the complementary expression, the average as weighted by the y-component of the wavevector, which is the same in each region. After substitution of $k_x = n_1\omega/c$ and $\kappa = in_2\omega/c$, $\tau_{\gamma,\text{meas}}$ bears a striking resemblance

to equation (5.42). This is expected, as the Goos-Hänchen shift dominates the total photonic delay τ_γ in the tunneling regime but becomes negligible in the Fabry-Perot regime.

Equations (6.2)-(6.4) suggest that the double-prism system is incapable of directly measuring delay variations caused by the Goos-Hänchen shift. Any delay incurred through the k_y component is identically offset by a glass propagation reduction of equivalent magnitude. Note that this is true for any value of the prism apex angle; for these examples we have used equilateral prisms, but the same result is obtained for right-angle prisms, rhombi, or any other arbitrary triangular or quadrilateral shape. And while we have only shown normal incidence at the exterior prism faces, the result is the same for non-normal incidence, as the boundaries of the system must always be defined by planes normal to the input and output wavevectors.

There is a considerable practical significance to this observation, as the Goos-Hänchen shift is responsible for the majority of the predicted delay τ_γ. For the devices discussed in the next section, τ_γ may be on the order of picoseconds at a Fabry-Perot resonance, but $\tau_{\gamma,\mathrm{meas}}$ is smaller by a factor of 100. Similar discrepancies can be observed in the barrier region.

Rather than consider this a failing of the design, we could perhaps consider this a "feature." By eliminating the ability to measure variations that occur in the y-dimension, we are able to isolate the variations that occur due to the x-direction. One could argue that such a measurement allows us to more directly probe the "tunneling" aspects of FTIR, and thus more intimately observe its relationship to the one-dimensional problem.

This practical issue has been overlooked by the majority of the literature [116–118, 122]. Haibel and Nimtz appear to have mentioned this effect in passing, stating that "The measured time was obtained by properly taking into consideration the beam's path in the prism" [3]. However, they make no further mention of the fact and do not provide experimental delay data for further scrutiny. In addition, they state that

their experiments were performed under conditions where the Goos-Hänchen shift approaches its asymptotic value, which suggests that their experimental measurements should be identically zero. This is a curious omission given their conclusion that the observed delay was entirely due to the Goos-Hänchen contribution.

It should be noted here that this separation into "measurable" and "unmeasurable" parts is specific to a time-domain measurement scheme, though it should be applicable to any direct measurement of time delay in such a structure. The FTIR phenomenon provides other observable quantities that can be exploited to make indirect measurements that are correlated to delay values according to theory. Of particular note here are two different measurements performed by Deutsch and Golub [123] and Balcou and Dutriaux [95].

In [123], the authors reported the measurement of the optical analog to the Larmor clock treatment proposed by Büttiker [110]. In this experiment, they draw parallels between the spin precession of the electron in a Larmor clock measurement and the polarization rotation of photons tunneling through a birefringent material in FTIR. By measuring the Stokes parameters of the tunneled beam as a function of incidence angle, they were able to infer a Larmor-like tunneling delay and achieved reasonably good agreement between theory and experiment. However, they note that the strong anisotropy in their system prevents them from unambiguously determining the Larmor time as a traversal time, and that it is still an open question whether the Büttiker treatment, which gives a complex time delay, has a significant physical meaning [90, 124].

The work described in [95] instead focuses on the Goos-Hänchen shift Δy and the deviation in output angle caused by the selective frequency transmission of FTIR, both of which they measure precisely. From these position and angle measurements, they infer the tunneling delay times in both reflection and transmission. Their measurement of Δy seems to confirm the "phase time" predictions of Hartman and others, while the measurement of deviation of output angle agrees with the "loss time"

suggested by Büttiker's treatment. They also observed that the "phase time" was symmetric in transmission and reflection but depended on the boundary conditions (TE- vs. TM-polarized input beams), while the "loss time" was heavily asymmetric in transmission and reflection but showed no dependence on the boundary conditions. They contend that this makes the "loss time" the more relevant measure of time spent within the barrier, though this conclusion overlooks some significant criticisms of this formulation [90].

In addition, it is our feeling that their conclusion overlooks the fact that boundary conditions do have a significant effect on the time spent in the barrier region, as a different boundary condition corresponds to a different effective barrier height and smaller wavefunction presence in the barrier region. Much of this is clearly seen in the derivations performed in the previous chapter. It should be no more surprising that the "phase time" measurement saturates to a different value for TE- and TM-polarized photons than that it saturates to different values for different barrier heights.

The dichotomy between "failing" and "feature" in our experiment is perhaps a bit clearer given these results. We cannot measure the true time spent in the barrier region, as a significant portion is tied up in the y-direction mechanics that are unavailable to us. Based on that, one might question the relevance of any measurement that can be performed in this configuration. However, by eliminating that component, we are free to make precise measurements of the portion $\partial\phi_0/\partial\omega$, which describes the evanescent x-component of the field and captures the aspects of the process that are specific to tunneling, making them potentially more interesting than methods that are incapable of separating the two.

6.2 Tunable double-prism structure

While it would be ideal to build a double-prism structure as proposed by Steinberg [82], there are practical concerns that make it difficult to achieve experimentally.

A glass-air-glass FTIR structure has a large index contrast and leads to very short interaction distances on the order of $\lambda/2$. Achieving a uniform air gap of less than 400 nanometers is not an easy feat outside of a cleanroom environment, especially for large surface areas. Initial attempts at achieving a tunable air gap through mechanical means lead to significant stress-induced refractive index changes that would have been an order of magnitude larger than the tunneling delays we desired to measure.

However, the tunneling process is governed by κL, the barrier "opacity." Thus, instead of increasing the barrier length, we can consider increasing the "height" by changing κ. We have chosen this approach for our experiment, modulating the barrier height by introducing a liquid crystal into the barrier region. The liquid crystal (LC) can be tuned by an externally-applied voltage to change the index contrast of the glass-LC interface and subsequently the effective barrier height. Our system has the added advantage of having no moving parts and introducing no significant beam deviations during the tuning process, both of which are liabilities inherent to the mechanically-tuned version initially proposed.

Our experimental implementation of a double-prism system is shown in figure 6.3. A liquid crystal cell is sandwiched between two equilateral prisms made of high-index N-SF11 glass, which has a refractive index of $n = 1.77$ at $\lambda = 727$ nm [125]. A small amount of index-matching fluid ($n = 1.70$) is applied to the prism-cell interfaces to minimize unwanted reflections. The high-index glass is necessary to create a large enough index contrast between liquid crystal and substrate to observe FTIR.

Liquid Crystal cell construction is a multi-stage process that occurs in a cleanroom environment. Two 3-mm thick N-SF11 substrates are sputter-coated with a 30-nm layer of Indium Tin Oxide (ITO) to create a transparent conductive layer that can serve as an electrode. Next, a 15-nm alignment layer of polyimide is spin-coated on top of the ITO layer and thermally cured. The substrate is then subjected to a mechanical buffing technique in which a rotating felt drum passes over the polyimide layer. This process shears the polyimide layer, which will cause the liquid crystal to

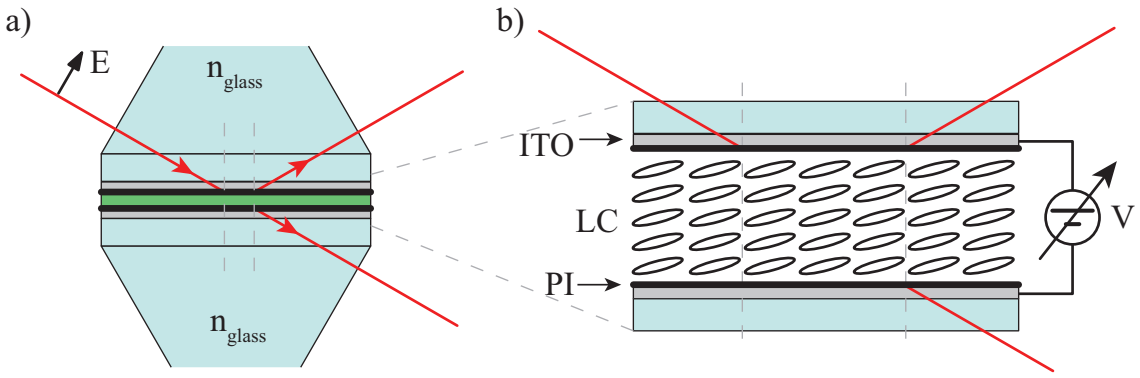

Figure 6.3: Diagram of the double-prism system. In (a) a liquid crystal cell is sandwiched between two glass equilateral prisms. An exploded diagram of the liquid crystal shell is shown in (b), with the liquid crystal (LC), Indium Tin Oxide (ITO) layer and polyimide alignment (PI) layer labeled.

align along the buffing direction when the cell is filled [126].

Epoxy spacers are applied to four spots on one substrate and the substrates are pressed together, processed surfaces facing one another. The two substrates are offset from each other slightly so that the ITO electrodes are accessible after the cell is complete, as illustrated in Figure 6.4. Pressure is applied while the epoxy cures, which creates a void of approximately 12 μm between the substrates. We have performed this procedure several times with low concentrations of microspheres mixed into the epoxy, and despite varying the diameter of the spheres from 3 to 10 μm the void was consistently 12 μm, indicating that this thickness is characteristic to the pressure applied and the epoxy volume or viscosity. Measurements of the cell thickness are made with a spectrophotometer before proceeding further.

The cell is then filled with liquid crystal solution through capillary action. The liquid crystal chosen was E7, a uniaxial nematic liquid crystal which is popular, commercially available, and inexpensive. The edges of the cell are sealed with epoxy to prevent evaporative loss and deterioration of the liquid crystal, as well as to prevent foreign objects from entering the cell. Wires are attached to the exposed ITO sections with conductive silver epoxy to accommodate an external voltage source.

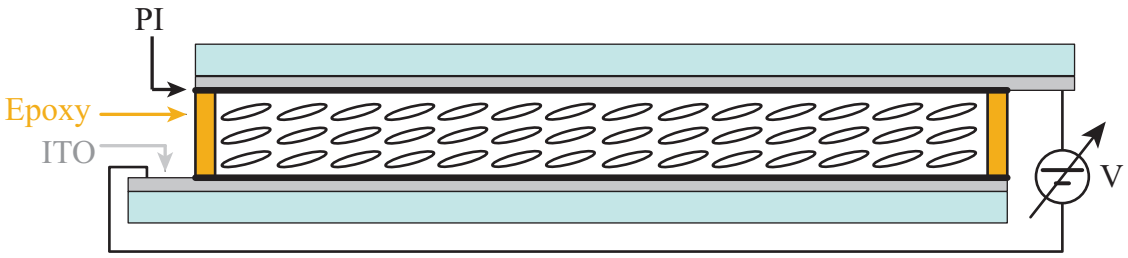

Figure 6.4: Close-up diagram of cell construction. Two high-index N-SF11 glass substrates are coated with an Indium Tin Oxide (ITO) layer and a Polyimide (PI) alignment layer. The cell is created by sandwiching epoxy spacers in-between the substrates and applying pressure. Once the epoxy cures, the liquid crystal is added from the side through capillary action.

While an in-depth discussion of liquid crystals is beyond the scope of this thesis, we will briefly describe the characteristics which are essential to our experiment. A liquid crystal has many of the properties of a liquid, such as flow and capillary action, but the individual molecules align with one another as in a crystalline structure. We will limit our discussion to uniaxial nematic liquid crystals, whose molecules have a single orientation axis and can be conceptually thought of as cylindrical rods. The asymmetry of the liquid crystal leads to birefringence, such that light polarized parallel to the liquid crystal axis experiences an "extraordinary" refractive index n_e and light propagating along the axis of the liquid crystal experiences an "ordinary" refractive index n_o regardless of polarization. The orientation of these molecules can be changed by the application of an external electric field, with the long axis of the cylinder tending to align to the electric field vector. Thus, the optical properties of the liquid crystal can be electronically controlled by adjusting the angle between the polarization vector of incident light and the orientation axis or "director" of the liquid crystal [127]. For E7, the extraordinary and ordinary refractive indices are $n_e = 1.718$ and $n_o = 1.514$ at $\lambda = 727$ nm [128].

The buffing process performed on the alignment layer of polyimide in our cell causes the liquid crystal to naturally orient along the direction of buffing, perpendic-

ular to the surface normal. By applying a voltage to the ITO electrodes, we impose an electric field that supplies a torque to the crystals and attempts to twist their orientation. If a sufficient voltage is applied the crystal will twist a full 90° and its director axis will align to the electric field vector, which is parallel to the surface normal.

If we consider the cell alone in the absence of the prisms or applied voltage, it is clear that light incident normal to the surface will experience an index n_e if it is polarized along the liquid crystal director and n_o if it is perpendicular to the director. Applying a voltage to the cell will continuously change the index seen by the parallel polarization from n_e at $V_{\mathrm{LC}} = 0$ to n_o at $V_{\mathrm{LC}} = V_{90°}$, while the perpendicular polarization will experience no change. The dependence of liquid crystal rotation angle on applied voltage can be determined experimentally from birefringence data obtained through transmission measurements on each polarization as described in [129].

Measurements of FTIR must be performed at oblique incidence, of course. At high incidence angles, tuning of the liquid crystal angle is equivalent to tuning the critical angle of the glass-LC interface. The relevant indices of E7 and N-SF11 dictate a range of achievable critical angles from 58.5° to 76.1°. Thus, the desired angle of incidence is between 60° and 65° such that we can make measurements in both the FTIR and Fabry-Perot regimes by tuning the critical angle of the structure. The equilateral prisms are necessary as a coupling aid to accomplish this, otherwise we would be unable to achieve the appropriate angle in the glass region.

6.3 Matrix model

The theory of plane wave propagation through anisotropic materials has been developed over the course of more than a century, as has the analysis of refraction and reflection at interfaces between isotropic and anisotropic materials [130]. Several

papers in the early 1970's addressed the particular problem of propagation through arbitrarily-oriented anisotropic materials using a 4x4 matrix technique [131–133]. In this technique, Maxwell's equations are solved in matrix form for two dimensions of the field (usually E_x, E_y, H_x, and H_y). While complicated, this general treatment is capable of finding solutions even when the anisotropy of the material varies in one dimension, as is the case in cholesteric liquid crystals. For the slightly simpler problem of a uniform anisotropic medium, this method has been used to generate analytical solutions for certain special cases of the anisotropy orientation [130, 134, 135].

We have chosen to employ this matrix technique to numerically calculate the expected delay in our double-prism system. Since this technique solves Maxwell's equations for the system, it produces valid solutions even under conditions which result in FTIR. Our implementation is virtually identical to that given in [133]. We calculate the complex transmission and reflection coefficients as a function of crystal orientation for a particular frequency ω and incidence angle θ. The crystal orientation is constrained to lie in the plane of incidence, and is described by the angle ϕ between the crystal axis and the surface normal. Repeating the calculation at different input angles θ and photon frequencies ω gives us the necessary information to calculate the frequency and angular derivatives of the transmission and reflection phases. This gives us a complete picture of the evolution of the system as the liquid crystal is rotated from $\phi = 0°$ (perpendicular to the interface) to $\phi = 90°$ or $\phi = -90°$ (parallel to interface). Note that the limitation of incidence-plane rotation is imposed by our experimental configuration; the matrix technique is also capable of handling rotations about an arbitrary axis if desired.

Since this analysis only applies to plane waves, we need to modify it slightly to account for the spread of k-vectors present in our Gaussian beam. We have chosen to convolve the calculated phase of the transmission and reflection coefficients with a Gaussian function to simulate these effects. In addition, any disorder or irregularity in the liquid crystal alignment can be encompassed by this operation. This step is

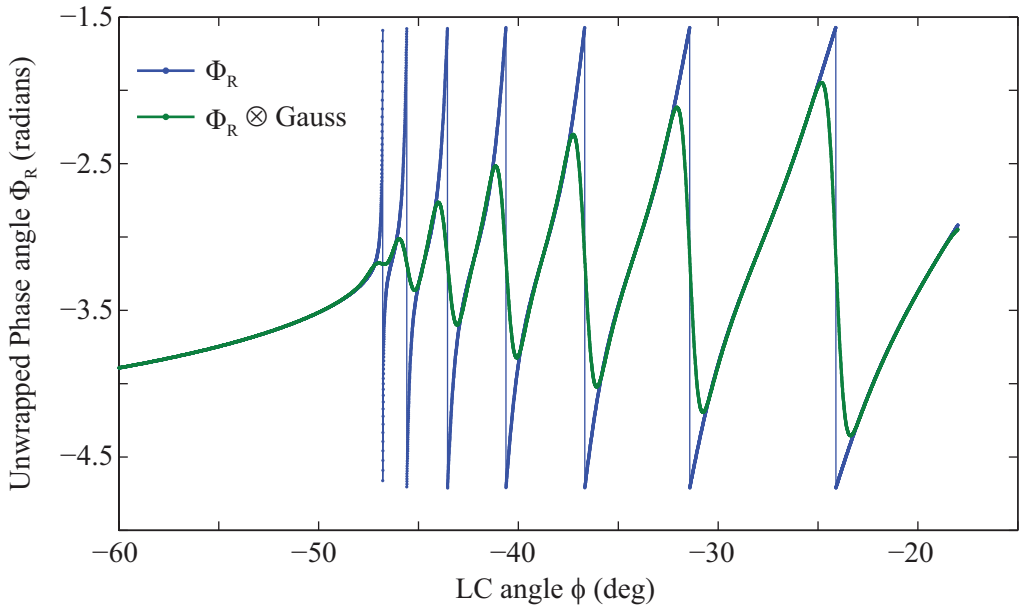

Figure 6.5: Unwrapped reflection phase Φ_R as a function of liquid crystal rotation angle ϕ. The blue curve is the phase output of the 4x4 matrix method simulation, which contains several π phase discontinuities at the Fabry-Perot resonances. The green curve is the convolution of that phase with a narrow Gaussian ($\sigma_{\text{Gauss}} = 0.35°$) to account for the effects of focusing. The indices of refraction used are $n_1 = n_2 = n_{\text{N-SF11}} = 1.76954$, $n_e = 1.633$, and $n_o = 1.54$, and the liquid crystal region was $L = 8\mu$m thick.

illustrated in Figure 6.5, which shows the unwrapped reflection phase Φ_R as a function of liquid crystal rotation angle ϕ. The blue curve is the raw output of the simulation for a plane wave, and contains π phase discontinuities at each Fabry-Perot resonance. The green curve is phase after the convolution has been performed. The width of the Gaussian was chosen as $\sigma_{\text{Gauss}} = 0.35°$ based on our experimental parameters. Unlike Φ_R, the transmission phase Φ_T (not shown) is continuous after unwrapping and does not contain regions of inverted phase slope. The unwrapping process is required to remove a number of 2π phase discontinuities that do not have physical significance, unlike the π phase shifts in Φ_R, but would otherwise impact the convolution process.

The discontinuities present in Φ_R lead to a peculiar effect in the predicted delay $\partial\Phi/\partial\omega$, as shown in Figure 6.6. The two delay predictions are similar in the off-

resonance regions, but the reflection delay exhibits sharp dips corresponding to the regions of inverted phase slope. These abrupt dips occur on every Fabry-Perot peak, where we would instead expect a delay *maximum* because of the extra time spent in the cavity region. This is due to the π phase discontinuity that occurs in reflection at the Fabry-Perot peaks [136]. This effect would not occur for a plane wave because it contains a single k-vector component. However, our focused beam contains a spread of k-vectors from either side of the phase discontinuity, and their interference causes this anomalous effect. The width of the Gaussian convolution function determines the size and depth of these features; as the Gaussian function narrows and approaches a Delta function, the reflection delay begins to look more and more like the transmission delay, but with a spike discontinuity at the Fabry-Perot peak.

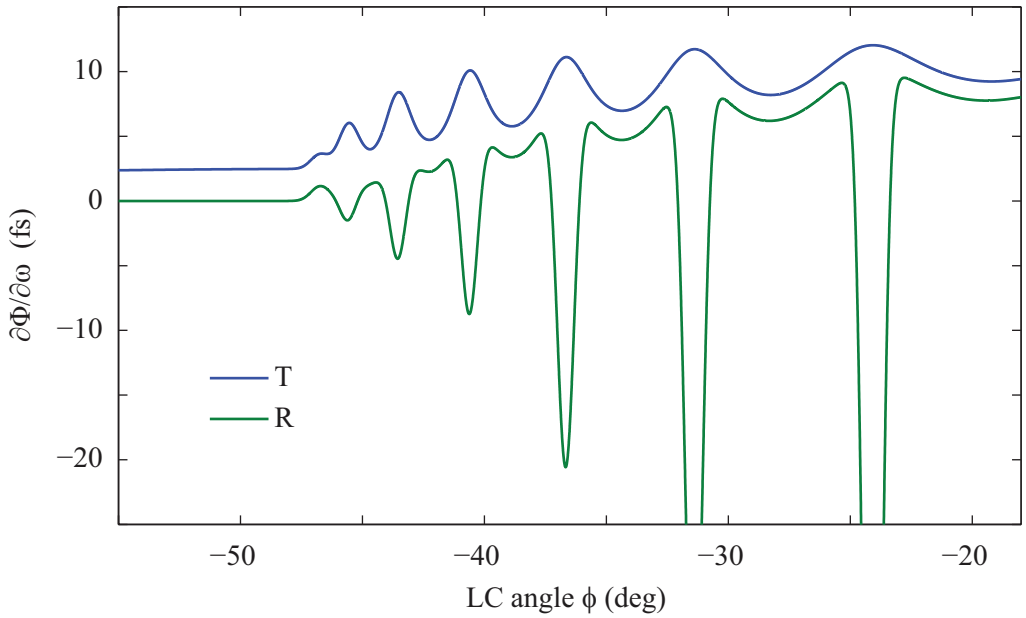

Figure 6.6: Transmission and Reflection delays $\partial\Phi_T/\partial\omega$ and $\partial\Phi_R/\partial\omega$ as functions of liquid crystal angle ϕ. The parameters used here are the same as in Figure 6.5.

The transmission delay values are approximately 2.5 fs higher than those of the reflection curve, including a non-zero delay prediction in the tunneling region. Interestingly enough, the exact opposite is also present in the $\partial\Phi/\partial\theta$ terms, such that the

total delay τ_γ for the reflection and transmission processes are identical. It is noteworthy that the delay offset is identical to $L(n_e - n_o)/c$, suggesting that the liquid crystal asymmetry is the source of this effect. Furthermore, reversing the rotation direction of the liquid crystal flips the asymmetry, giving the reflection delay a 2.5-fs lead over the transmission delay. In test simulations where the liquid crystal is replaced with an isotropic medium with a varying index, the asymmetries disappear and all three of the quantities (τ_γ, $\partial\Phi/\partial\omega$, $\partial\Phi/\partial\theta$) are identical in transmission and reflection.

Our interpretation of this is that the total transmission and reflection delays ($\tau_{\gamma r}$ and $\tau_{\gamma t}$) are constrained to be identical by the symmetry of the glass-LC-glass barrier system, as would be expected by the cavity interpretation. However, the local asymmetry of the liquid crystal layer imparts a change in the distribution of that delay amongst the two phase derivative terms. This means that the LC asymmetry causes a different Goos-Hänchen shift at the two interfaces, which leads to a different stationary phase constraint on $\partial\Phi/\partial\omega$ for transmission and reflection. This is consistent with the effect disappearing when the barrier medium becomes isotropic, a feature that we also observe in simulation. In addition, TE-polarized light does not experience this reversal effect in our simulation, in agreement with theoretical predictions [137].

To complete the model we need to relate the crystal orientation ϕ to an applied voltage V. This is determined experimentally with the setup described in reference [129]. A laser and polarizer are arranged to provide linearly polarized light. The cell is placed such that the beam strikes it at normal incidence and the plane in which the liquid crystal axis rotates makes an angle of 45° with the polarizer. An analyzer is placed after the cell to isolate the components of the transmitted field that are parallel and perpendicular to the polarizer, and transmission measurements are made for both polarizations as a function of voltage applied to the liquid crystal cell. This data is processed and fit to determine the numerical birefringence curve $\Delta n(V)$. The birefringence Δn can then be related to ϕ using the well-known equations for

the angle-dependent refractive index of a birefringent material [34]. Combining these gives us numerical relations with which we can convert between ϕ and V in either direction.

The matrix calculation also calculates the Goos-Hänchen shift Δy and total delay τ_γ. An exhaustive analysis of these predictions is not necessary here, but it is worth briefly making note of their approximate values to demonstrate that the predicted τ_γ values are significantly larger than those for $\tau_{\gamma,\text{meas}}$. The Goos-Hänchen shift varies from approximately 2 μm deep in the barrier region to hundreds of microns in the Fabry-Perot region. The group delay τ_γ reflects these swings, varying between 10-30 fs within the barrier region and rising to over 100 fs near the Fabry-Perot peaks. These discrepancies are very large, and the experimental measurements presented later in this chapter will confirm that $\tau_{\gamma,\text{meas}}$, not τ_γ, is the quantity of interest in such an experiment.

6.4 The Hong-Ou-Mandel effect

The time delays we expect to observe in this structure are on the order of tens of femtoseconds, necessitating the use of a very precise measurement system for the observed delays. For this purpose we chose to employ a Hong-Ou-Mandel (HOM) interferometer, which exploits quantum interference effects to achieve sensitive time measurements [138, 139]. We will now briefly review the operating principles of a HOM interferometer.

Figure 6.7 shows a simple HOM interferometer setup. A pump laser at frequency 2ω incident on a parametric down-conversion (PDC) crystal generates entangled pairs of "signal" and "idler" photons at frequencies $\omega_s = \omega + \delta\omega$ and $\omega_i = \omega - \delta\omega$, such that $\omega_s + \omega_i = 2\omega$. Signal and idler photons travel through different arms before being brought together at a perfect beamsplitter. The two output ports of the beamsplitter are sent through bandpass interference filters before collection by single-photon

detectors. The detector outputs are then post-processed with circuitry that tracks the number of "coincidences," events when both detectors fire within a fixed time window. This coincidence window is necessarily very large compared to any other time scale in the system [140, 141], so any two detection events caused by a single down-conversion event will always fall within this window.

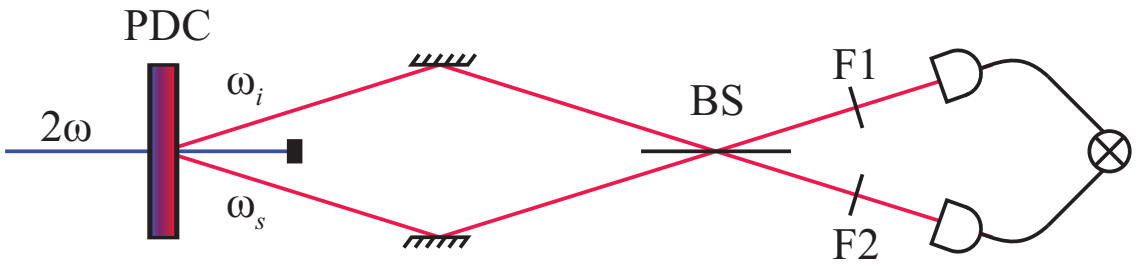

Figure 6.7: Basic Hong-Ou-Mandel interferometer. A parametric downconversion crystal (PDC) generates entangled pairs of signal and idler photons which are brought back together on a beamsplitter (BS) before arriving at two detectors. When the interferometer is balanced, quantum path interference leads to destructive interference that reduces the rate of coincidences detected at the detectors.

When the signal and idler photons arrive at the beamsplitter each has an equal probability of being transmitted or reflected, leading to four possible outcomes: both photons transmitted (TT), both photons reflected (RR), and two combinations with one photon reflected and one transmitted (TR, RT). The circuitry will register a coincidence for the TT and RR events as long as the photons arrive within the fixed time window. However, they will not fire for the TR or RT events, as in those two situations both photons arrive at the same detector. By measuring coincidences, the system is postselecting the subset of events where both photons arrive at different detectors.

If the interferometer is heavily unbalanced, such that the difference in photon arrival time at the beamsplitter is significantly larger than the coherence length of the signal-idler field [142], quantum interference cannot take place and each of the four outcomes is equally likely. However, if the two arms are balanced and the photons arrive at the beamsplitter simultaneously, quantum-mechanical destructive interference

occurs between the TT and RR results, and no coincidence counts will be measured by the detection system. If the path length asymmetry of the interferometer is scanned through the balance point, either by moving the beamsplitter itself or by introducing a variable-path-delay element in one arm, the coincidence rate will exhibit a "Hong-Ou-Mandel dip," continuously changing from no coincidences at the balance point to the background coincidence rate on either side of that point.

In practice, the visibility of the HOM dip is not one-hundred percent. Accidental coincidences due to dark counts, stray light, simultaneous downconversion events, asymmetry in the beamsplitter, and imperfect optics will all reduce the visibility. However, the visibility of the dip is not closely tied to timing precision unless it is so low that the dip becomes unrecognizable compared to noise-level fluctuations.

The time resolution of this interferometer is primarily determined by the width of the HOM dip. This width is related to the frequency weight function of the downconverted photons as dictated by the crystal and pump beam parameters in the interaction Hamiltonian [143]. The width of the HOM dip is proportional to the Fourier transform of this weight function. If we assume the weight function is roughly Gaussian with bandwidth $\Delta\omega$, the HOM dip width is then proportional to the inverse of the down-conversion bandwidth, or $1/\Delta\omega$. Thus, broader down-conversion bandwidths give narrower HOM dips and better time resolution. However, in most setups, the bandwidth and transmission function of the interference filters acts as a more stringent limitation than the down-conversion process.

In the seminal paper on this effect, Hong, Ou, and Mandel demonstrated a dip with visibility in excess of 80% and a width of about 16 μm of beamsplitter movement, or 100 fs of path length [138]. They noted that the time resolution of this system could be better than 1 fs. Later measurements by Steinberg et al. demonstrated resolutions of 4 fs in measurements of single-photon propagation velocity in glass [140] and 0.2 fs in measurements of the single-photon tunneling time in a one-dimensional photonic band-gap material [99] with dip widths of 35 to 50 fs.

6.5 Experimental setup

The experimental setup is shown in abbreviated form in Figure 6.8a. An ultraviolet pump beam incident on a nonlinear crystal generates spontaneous parametric down-conversion (SPDC), providing a source of entangled photon pairs. The photon in the upper arm is incident on the double-prism FTIR system in either a reflection or transmission geometry. The lower arm contains a "trombone" system constructed from a corner cube retroreflector mounted on a motorized linear translation stage. The two arms of the interferometer are then combined at a non-polarizing 50/50 beamsplitter to create a HOM interferometer. Beamsplitter outputs are coupled into single-mode fiber and sent to avalanche photodiode single-photon counting modules, the output of which is sent to coincidence counting circuitry.

Inset (b) in Figure 6.8 shows a schematic diagram of the interaction region. The liquid crystals have an effective index n_{eff} determined by their rotation angle, which is controlled with the applied voltage V. The interfaces between n_{glass} and n_{eff} serve as the boundaries of the barrier region. Reflected light undergoes a Goos-Hänchen shift as indicated by the dotted vertical lines. Inset (c) shows an example data trace taken with the coincidence-counting circuitry, which we then digitize and fit in a computer.

A more complete experimental diagram is shown in Figure 6.9. The pump source is a continuous-wave Coherent Innova Sabre argon-ion laser operating at 364 nm, producing up to 1 W of power with a linewidth of approximately 3 GHz. This laser pumps a 3-mm BBO crystal to generate entangled photons at 727 nm by Type I non-collinear SPDC. The crystal is aligned and angle-tuned such that the generated photons are emitted in a cone with a half-angle of approximately 0.1 radians, or a little under 6°. Our particular crystal has a cut angle of 29.2° for optimized second-harmonic generation at 800 nm, so this alignment process includes an approximately 3° rotation from the pump normal to achieve phase matching for PDC at 727 nm (32.3°). The pump beam is focused on the crystal to increase the intensity and subsequently the SPDC generation rate. This also increases collection rates, as the

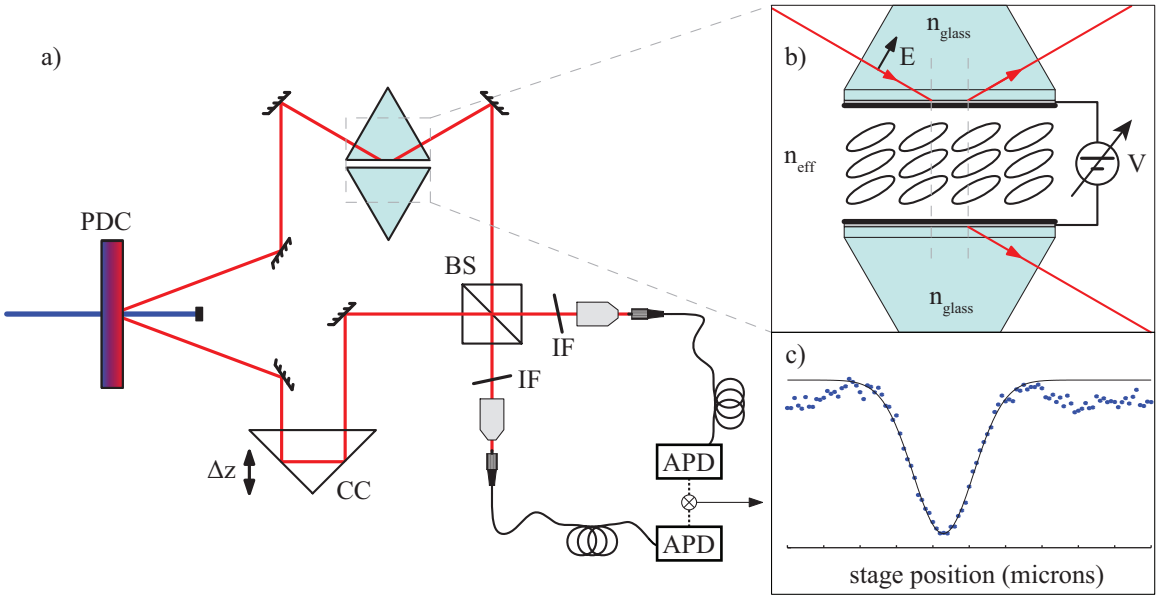

Figure 6.8: Abbreviated setup diagram showing FTIR measurements in the reflection geometry. Panel (a) shows the Hong-Ou-Mandel arrangement and placement of components. PDC is the parametric downconversion crystal, CC is a corner cube retroreflector, BS is a non-polarizing 50/50 beamsplitter, IF are 10-nm bandpass interference filters centered at 727 nm, APD are avalanche photodiode single-photon counting modules. Inset (b) shows a detailed view of the FTIR region, with two prisms of index n_{glass} and a barrier region of liquid crystals with an effective index n_{eff} due to the applied voltage V. Inset (c) shows an example data trace containing a Hong-Ou-Mandel dip, along with a numerical Gaussian fit.

single-mode fibers in our detection system can only collect photons generated from a relatively small active area on the crystal.

We use Type I down-conversion for this experiment because the two generated photons have the same polarization state. Our pump is horizontally polarized (in the plane of the diagram), generating down-converted photons that are vertically polarized (perpendicular to the plane of the diagram). We chose this over Type II down-conversion, in which the two photons have different polarization states, to avoid several potential pitfalls. The most obvious issue is that since the photon in the upper arm could be in either polarization state, we would have both TE and TM polarization states incident on our prism system. We could eliminate one or the other

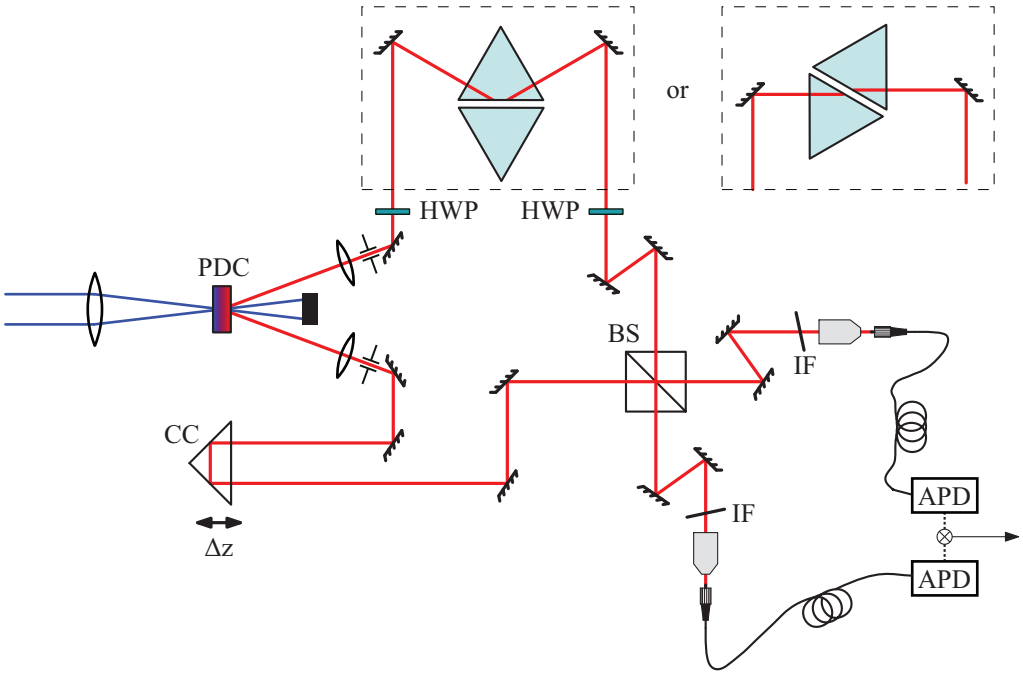

Figure 6.9: Complete setup diagram, including all optical elements. PDC is the parametric downconversion crystal, CC is a corner cube retroreflector, HWP are half-wave plates, BS is a non-polarizing 50/50 beamsplitter, IF are 10-nm bandpass interference filters centered at 727 nm, APD are avalanche photodiode single-photon counting modules.

by introducing a polarizer in the lower arm to perform pre-selection, but this would reduce our effective generation rate by half, which is not desirable.

Another potential issue with Type II down-conversion is that birefringence in the crystal causes a small relative delay between the two generated photons. This delay depends on the polarization of each and the exact position along the optic axis at which the generation took place. Experimentally this means that there is a minimum uncertainty in the relative delay between the two photons which manifests itself as a broadening of the Mandel dip and a decrease in timing resolution. The magnitude of this broadening depends on the crystal length, but for a 3 mm crystal it could be as large as 300-600 fs ($\Delta n \approx 0.066$). While we could observe a Mandel dip under such conditions, it would significantly reduce our timing accuracy. In combination with the reduced generation rate this makes Type II less desirable than Type I since the

latter avoids both of these problems.

The two down-converted photons then proceed through different arms of the interferometer. In each arm, a small lens of approximately 10-cm focal length is positioned about 20 cm from the down-conversion crystal. These lenses image the downconversion plane to a spot within each arm, which serves two purposes. The first is to ensure that the beam size remains small near the interaction region of the prism structure, which helps minimize the effects of large-scale irregularities in the prisms or liquid crystal cell. The second is to prevent the mode size from growing so large that it overfills the microscope objectives in the collection arms. This is a significant improvement in collection efficiency in a long system such as this one; observed count rates improved by more than a factor of five when the lenses were inserted and aligned properly.

Each arm also contains an iris immediately following the lens. These irises do not serve any purpose for data collection, but are extremely valuable for system alignment and troubleshooting. We will discuss this in more depth later in the chapter.

In the lower arm, the photon proceeds through two mirrors to encounter a "trombone" system consisting of a corner cube retroreflector mounted on a motorized linear translation stage (Aerotech model ATS50-25-M-2). This system introduces a controllable amount of path delay for the photon in the lower arm. The stage is capable of step sizes as small as 50 nm and has a unidirectional repeatability of 300 nm. The resolution far exceeds the motion intervals of 1 μm that were used in the experiments, but the unidirectional repeatability was critical for achieving consistency between successive data traces. Since our trombone arm introduces a double-pass path delay, an error of 300 nm in position would cause a 2 fs error in the location of the Mandel dip minimum. In our experiments the observed error at high count rates was closer to 1 fs, suggesting that the repeatability error distribution function was a little narrower than 300 nm. In any event, we believe that this source of error established the lower bound on our system's time resolution. Coincidence count rates

and thermal stability of the system dictated how close our data would come to this lower bound.

We chose to use a corner cube instead of a more traditional right-angle prism for greater alignment fidelity. We found that the right-angle prisms available in our laboratory produced a vertical deviation in the output wavevector. Since this error was inherent to the prisms, it led to alignment instability as the stage was translated from one limit to the other. The corner cube retroreflector did not exhibit this behavior, which led to much more stable alignment across the entirety of the translation stage's 2.54-cm travel length. The two pairs of mirrors before and after the corner cube allow for precise parallel alignment of the input beam path to the translation stage's motion vector. A small angle between these two vectors causes a lateral shift of the output beam as the stage advances, which leads to mode mismatch and a reduction in the visibility of the quantum interference. For even small angles, this can cause the Mandel dip to disappear, and for more serious alignment errors a loss of coincidence can occur. These factors make the control afforded by these four mirrors critical to stable operation.

In the upper arm, the photon interacts with the double-prism test system. Figure 6.9 shows the two possible configurations under which the system can be tested. The diagram demonstrates the reflection geometry, in which photons reflected by the barrier are collected and measured. The inset shows the transmission geometry, in which photons that tunnel through the barrier and emerge on the other side are measured. The prism system itself is the same for both measurements, but the mirrors immediately preceding and following the system must be adjusted to account for the change in orientation, necessitating a significant degree of re-alignment in the test arm when changing from one geometry to the other.

The two arms of the interferometer are then brought back together at a non-polarizing 50/50 beamsplitter to create the HOM interferometer. Beamsplitter outputs are passed through interference filters to suppress stray light from the pump,

pump fluorescence, and ambient light. The interference filters are 10-nm band-pass filters centered at 730 nm, which are then tilted 12° to re-center the bandpass region at 727 nm. The transmitted light is then coupled into a fiber-based collection module with another pair of mirrors. The collection module consists of a microscope objective on an x-y translation mount and a single-mode fiber (ThorLabs P1-630A-FC) mounted in an FC connector plate on a z-direction translation mount. These mounts are supported by a home-built ThorLabs cage substructure that mimics their KT110 free space fiber coupler system.

The fibers transport the light to the avalanche photodiode (APD) single-photon counting modules (PerkinElmer SPCM-AQR-14-FC). These modules have very low dark count rates (less than 100 counts per second) and a quantum efficiency of approximately 70% at 727 nm. They produce 30-ns long TTL output pluses upon a detection event, though there is a 50-ns "dead time" between pulses which limits the maximum count rates to around 10 million counts per second [144]. Our single-detector count rates rarely exceeded 100 thousand counts per second, so this limitation is of little consequence to our experiment.

A coincidence circuit takes the electrical output of the APDs and performs an array of logic functions, tracking the number of individual ("singles") events at each detector as well as the number of coincidence events. The circuit is home-built and based on a design published by Mark Beck [145] which is freely available online [146]. This particular model has since been retired in favor of newer models based on field programmable gate arrays (FPGAs). The new models connect to the computer by USB or RS323 cables and do not require the National Instruments PCI-6602 counter/timer board that our model does, which make them cheaper and easier to interface with. The discrimination time window of the circuit is nominally about 12-15 nanoseconds, which is sufficiently large to guarantee that any generated pair of entangled photons which successfully triggers both APDs will be counted. The width of this window also ensures that odd-order dispersion effects (including group

velocity dispersion) cancel out and do not artificially broaden the HOM dip [141]. A shorter discrimination window would result in better count statistics by reducing the number of random coincidence events, but may lead to dispersion-broadening of the dip and a reduction in timing accuracy.

6.6 System construction and alignment

System construction and alignment is a complicated, multi-stage process. Initially, down-conversion must be achieved in a sparse system as shown in Figure 6.10a. Irises I1-I4 define the two down-conversion beam paths and I5-I6 the pump beam path, all of which intersect in the crystal. The pump beam is aligned to the pump irises with two mirrors (not shown), and the collection modules are "back-aligned" to the down-conversion beam paths using four mirrors. This is done by sending the output of a Ti:Sapphire laser operating in continuous-wave mode at 727 nm "backwards" through the fiber output of one collection module and aligning the beam to the two appropriate irises. The process is repeated for the second module and pair of irises. Once this is complete, the back-alignment laser is disconnected and the collection modules are connected to the APDs. The pump beam and crystal angles are then adjusted until down-conversion is achieved and optimized.

Once down-conversion is optimized the translation stage and corner cube retroreflector are introduced to the lower arm as in Figure 6.10b. It is crucial that the wavevector of the down-converted light incident on the retroreflector is parallel to the translation stage motion axis to prevent beam walk-off. To accommodate this, the system is "front-aligned" by sending a continuous-wave 632-nm HeNe beam forward through the lower beam path by aligning it to irises I1-I2. The output of the retroreflector is allowed to propagate for approximately 5 meters to a screen upon which the beam center is marked. Mirrors M1 and M2 are then adjusted as the stage is translated between its two limits until the spot on the screen no longer moves. M1

and M2 are then considered "fixed" and excluded from subsequent alignment steps to prevent introduction of walk-off error.

The rest of the setup is straightforwardly constructed through a combination of back-alignment and singles/coincidence maximization. The beamsplitter, mirrors, and lenses are added to the setup as shown in Figure 6.9c. Mirrors M3-M9 are placed such that the estimated path difference between the two arms is less than 1 cm, while mirrors M10-M13 are placed to make the path difference between the collection modules and the beamsplitter faces approximately equal. The back-alignment beam is applied to the lower collection module, and mirrors M10-M11 are adjusted until the beam is approximately centered on the beamsplitter at normal incidence and operating height. Mirrors M3-M4 are adjusted to align this beam with irises I1-I2, and mirrors M5-M9 are adjusted until the upper arm is similarly aligned to irises I3-I4.

Next, the back-alignment beam is removed and an APD is connected to the lower collection module. Mirrors M3-M4 are adjusted to maximize singles counts while the upper arm is blocked, and M8-M9 are adjusted to maximize singles counts while the lower arm is blocked. Once that is complete, the back-alignment and singles maximization processes are repeated for the upper collection module, this time adjusting only mirrors M12-M13. Coincidences should be easily observed at this point, and fine-tuning of mirrors to maximize coincidence counts can take place. The particular mirrors used in this process depend on several different pieces of information including overall coincidence rate, the singles rate at each collection module due to each arm, and the efficiency of both APDs and collection modules.

We then introduce lenses L1-L3 to increase our coincidence count rate to more convenient levels. The focal length of all three lenses are approximately 10 cm. L1 is mounted on a three-dimensional translation stage and adjusted to focus the collimated pump beam on the crystal, while L2-L3 are mounted on two-dimensional (xy) translation stages. The back-alignment beam is reintroduced and the z-translation-stage

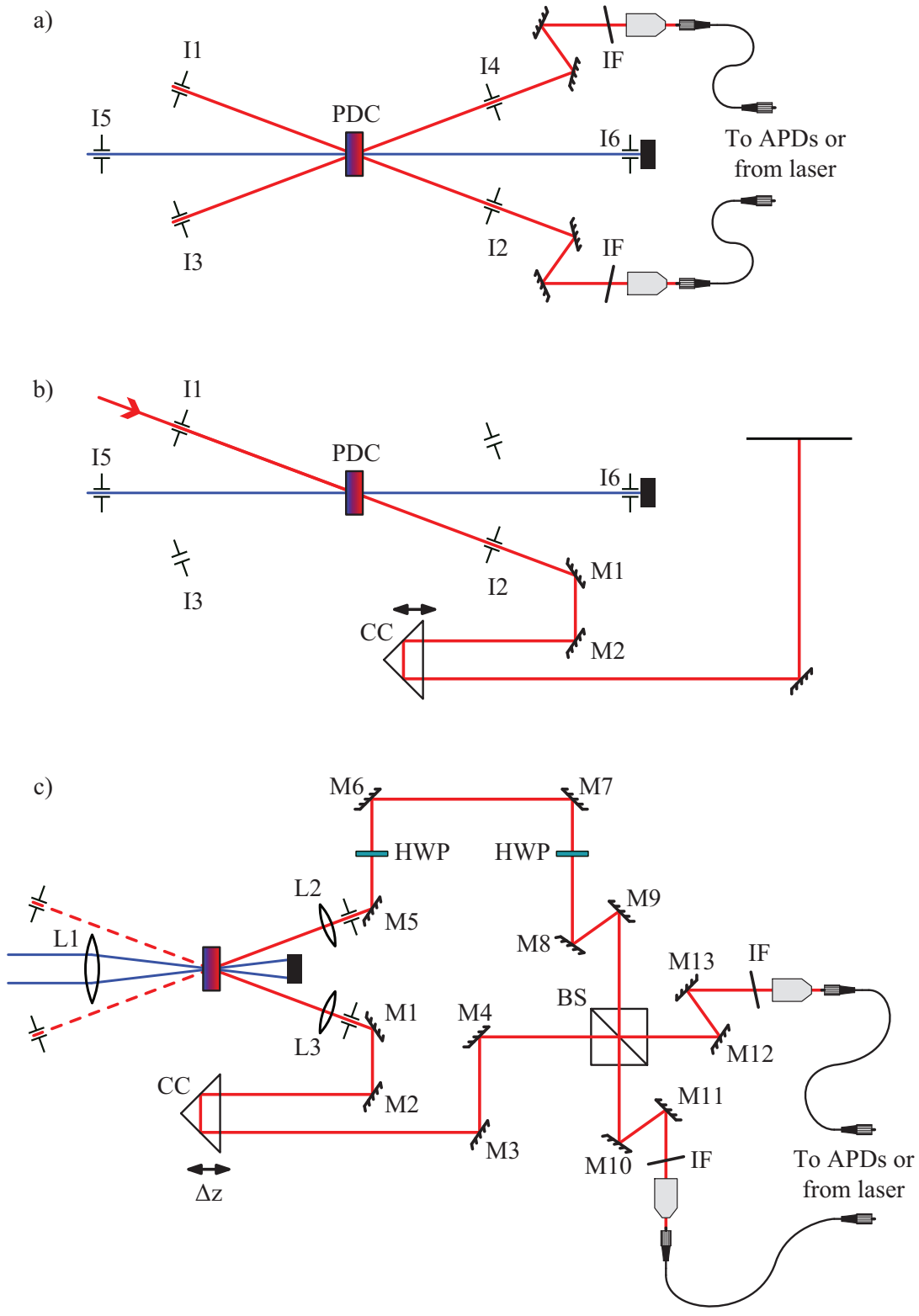

Figure 6.10: Illustration of the setup construction procedure. Down-conversion is aligned and optimized through a combination of back-alignment and photon counting as shown in (a). In (b) the motorized translation stage and corner cube are aligned. Panel (c) shows the final alignment stage for observing the Hong-Ou-Mandel dip. I1-I6 are irises, PDC the parametric down-conversion crystal, IF interference filters, M1-M13 mirrors, CC corner cube, BS beamsplitter, HWP half-wave plates.

of each collection module is adjusted until the microscope objective images the fiber tip to the test location of the upper arm. Lenses L2 and L3 are then hand-adjusted to re-image this point to the crystal. The distance between the crystal and each lens should be equal, though hand-adjustment limits the accuracy of this to roughly 1 mm. This error can manifest itself as a slight imbalance in the observed singles count rate when blocking either arm, as one arm will couple to the fiber more efficiently than the other if the imaging conditions of L2 and L3 are not identical. While it was not a significant problem in practice, the signal-to-noise may have been slightly better if these lenses were also mounted on xyz-translation stages. The addition of these three lenses can increase the observed coincidence rate very significantly; a rough estimate based on unrecorded observations places the contribution of these three lenses to the coincidence rate at a factor of around 10 or 20.

The APDs are then connected and coincidence is maximized again through mirrors M3-M4 and M6-M7. A coarse scan of the translation stage is performed to search for the Hong-Ou-Mandel dip, and once located fine scans are performed to assess its fidelity. Further mirror adjustment may be performed here to improve the visibility and shape of the dip, though it is not particularly necessary at this point.

Finally, the system to be tested is added to the upper arm. This usually necessitates a significant change in the positions of M6 and M7 because of the added glass path length. These mirrors are then used to back-align the upper arm to irises I3-I4, and adjusted again to maximize singles and coincidences under operating conditions. The coarse and fine scans of the translation stage are repeated to find the new location of the HOM dip and optimize its properties. Experimental data collection can then proceed as described in the next section.

6.7 Experimental procedure

On a conceptual level, data acquisition for these experiments is very straightforward once the system is constructed and aligned properly. The basic building block of the experiments is a HOM "trace," consisting of coincidence measurements made at sequential positions of the translation stage. In our traces, measurements are made over a 50-micron range of stage movement in 1-micron increments for a total of 51 data points per trace. At each new stage position we record the position, integration time, singles counts on each detector, and coincidence counts. The time and date of each trace are also recorded. In post-processing, the collected data is imported into MATLAB to fit each HOM trace with a Gaussian function as described in the next section. A position value that describes the minimum of the HOM dip is extracted from this fit. An individual experimental data run involves performing a large number of HOM traces with different voltages applied to the liquid crystal.

From the position values extracted in post-processing, we can pair these voltages with HOM dip positions to find the relationship between applied voltage and path delay. We can interchangeably talk about the "position" or "time" at which the HOM dip minimum occurs, so to avoid confusion we will refer to this value as the "centroid" of the HOM dip. The centroid can be expressed in either unit system as desired, with the understanding that a centroid position x corresponds to a centroid "time" $2x/c$ due to the double-pass nature of the trombone arm. Similarly, the difference in two centroid positions x_1 and x_2 is equivalent to the delay $\tau = 2(x_2 - x_1)/c$.

Since our experiment contains no absolute reference, all of our measurements are necessarily relative delays, or measured centroid differences between a reference voltage and a test voltage. As such, we cannot make statements about absolute delays or superluminality from our data. We can, however, characterize the delay curves as a function of voltage or liquid crystal angle and compare them to the theoretical predictions for our system to determine whether the model of tunneling is accurate.

In practice, there are a number of issues that complicate the data collection pro-

cess. The largest is thermal drift of the set-up and pump laser, as the temperature stability of the room housing the experimental setup is quite poor. It is not uncommon to observe temperature changes of up to 10° F over the course of a few hours. This leads to thermal expansion and contraction of the optical mounts, and can produce slow drifts in the relative path delay of the two arms.

A single HOM trace takes around one minute to acquire even when the coincidence rate is high, so this drift poses a significant problem. The drift error observed between two adjacent traces may not be very large, but the accumulated error over 10 or more traces frequently will be, especially for longer traces with higher integration times. Based on our empirical observations, the thermal drift effect is smaller than 1 fs of total change in the dip centroid value over approximately five minutes. However, over longer time scales the drift can exceed 10 fs per hour.

To combat this drift we employed a differential data acquisition method. We perform the a HOM trace at a reference voltage V_r, and then repeat the measurement at our test voltage V_t. An entire data set is thus a series of V_t traces interleaved with the same number of V_r traces. After post-processing provides us with the corresponding centroid positions Z_t and Z_r, we can calculate a set of values $z_i = Z_t^{(i)} - Z_r^{(i)}$ for every (V_r, V_t) pair that describe the drift-corrected centroid of the HOM dip.

An additional drift-related complication arises from the nature of tunneling itself. In the transmission geometry (see inset of Figure 6.9) very few of the photons incident upon the structure successfully make it to the detectors. This significantly reduces coincidence count rate, dip visibility, and fidelity, which subsequently reduces the accuracy of the Gaussian fitting algorithm. The transmission becomes so bad that some cases simply cannot be fit without increasing the integration time significantly.

However, increasing integration time also increases the likelihood and magnitude of drift errors. If we were to increase the integration time such that each test and reference trace took 30 minutes to acquire, the error with the differential acquisition technique could still be prohibitive. Instead, we chose to employ a "double-interleaving"

method in which we acquire data in smaller "chunks" and combine them in post-processing. To acquire M minutes of integration time at V_t we perform N differential measurements of duration M/N, and in post-processing combine the coincidence totals for each of the N individual test and reference measurements before attempting to fit the data.

As an example, if we want $M = 15$ minutes of integration time but want to limit our maximum single-trace time to 5 minutes to limit drift error, we perform $N = 3$ test traces of 5-minute duration at V_t interleaved with 3 reference traces of 5-minute duration at V_r. In other words, we perform 5-minute measurements for the voltage sequence $V = (V_t, V_r, V_t, V_r, V_t, V_r)$. The coincidence data for the odd traces are summed in the computer to give one compiled test trace with 15 minutes of integration time, and the reference trace is similarly compiled from the even traces. These compiled traces are then used in the fitting subroutine to determine one Z_t value for the voltage V_t and similarly for (V_r, Z_r).

Mathematically, this can be conceptualized as taking a collection of Gaussian functions $G(x, x_0, \sigma) = \exp(-(x - x_0)^2/2\sigma^2)$ with offsets x_0 and summing them. While the offsets x_0 are not truly random, they are small enough compared to σ that the resulting sum still strongly represents a Gaussian with slight broadening. Since the drift should be fairly consistent between any given pair of (V_t, V_r) measurements, the broadening and deformation characteristics of the sum should be strongly correlated. This means that both traces should show the same deformities and the fit algorithm should respond similarly to each. As a result, the broadening or deformities introduced through this method are suppressed in the final relative delay measurements as long as the signal-to-noise ratio is still high enough for the fit algorithm to give reasonable results.

The voltage V_r is usually chosen based on coincidence rate in order to maximize the accuracy of the reference trace. This often leads to situations where the coincidence rate at V_t is significantly lower than V_r, making the above "double-interleaving"

method somewhat inefficient, as it spends the same amount of time integrating at V_r as at V_t. In practice, we perform slightly more complicated interleaving operations to increase data collection efficiency.

To provide a specific example, let's say we want 30 minutes of integration time at V_t, but it only takes 5 minutes of integration time at V_r to get an acceptable reference trace. We could perform one 6-minute trace at V_t and one 1-minute trace at V_r, and then repeat that sequence a total of five times to get our 30 minutes of integration time. Alternatively, we might perform five measurement sequences of $(V_t, V_t, V_t, V_t, V_t, V_r)$ with 1 minute of integration time for each trace. The 6x1-minute method has some advantages over the 1x6-minute method; it performs some smoothing of short-term noise effects (1-minute duration or less) by spreading their effect across more data points within a single trace, and it keeps the mechanical motion of the stage independent of trace type.

For more severe imbalances, like 30 minutes for every V_t but only one minute for V_r, we oversample V_r in order to keep the time between reference traces reasonable, usually no more than 10-12 minutes. Sparser sampling limits our accuracy due to the slow thermal drifts that we've experimentally observed in the system, which take place on the order of 10-30 minutes.

Accumulation of this data is performed mostly unattended with the help of a group of Labview virtual instruments ("VIs") that we have written specifically for this HOM setup. The Labview code provides a graphical user interface within which one can set all of the relevant experimental parameters and initiate the scan sequence. This "master VI" then calls a variety of sub-VIs to handle portions of the experiment, including performing stage operations based on the scan parameters, querying the coincidence circuitry for accumulated APD data, applying voltages to the Liquid Crystal cell, constructing the interleaving scheme, and writing the data to raw text files for storage.

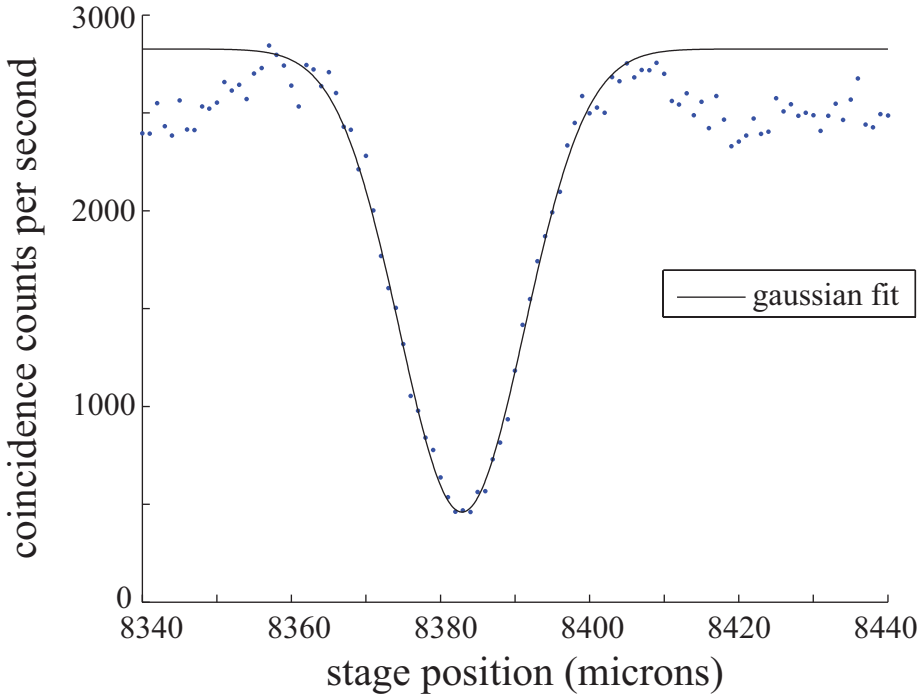

Figure 6.11: Sample Hong-Ou-Mandel trace along with Gaussian fit.

6.8 Data and post-processing

Once the coincidence data has been collected, it goes through several postprocessing steps to extract accurate values for the FWHM, position, and visibility of the dip for a given trace or set of traces. First the traces are imported into MATLAB from the raw text data files. The traces are combined according to the interleaving settings of the particular data run to create sequences of alternating V_t, V_r data trace pairs. Figure 6.11 shows an example data trace generated in this fashion, along with the accompanying functional fit. The small "humps" on either side of the dip are believed to be due to a photon bandwidth or mode match asymmetry between the two arms caused by the 3° rotation necessary for downconversion crystal alignment.

These sequences are then processed by the fitting module, which attempts to fit each individual trace using a nonlinear least squares method. The functional form

used for the fit is

$$y(x) = D - Ce^{-4\ln 2(\frac{x-B}{A})^2} \tag{6.6}$$

where A is the FWHM and B is the average position. While C is the maximum "depth" of the dip, it is clear from Figure 6.11 that D is not an accurate estimate of the background coincidence rate because of data weighting. When calculating the fit, the data points are subject to a weight function of the form

$$w(x) = e^{((x-x_0)/18)^4}. \tag{6.7}$$

The value of x_0 is chosen to be the location of the minimum coincidence value in the data trace as long as that point occurs within the center 50% of the domain. If it does not, x_0 is arbitrarily set to the mid-point of the x-axis. This super-Gaussian weight function emphasizes fitting of the dip region and reduces the influence of the "humps" and background, which become a significant source of error in low-transmission traces. The value of 18 was chosen empirically by visually assessing how well the fit matched moderately noisy data traces. Coincidentally, the average FWHM of the data traces is around 20 μm, which means that this weight function drops to zero at roughly the same point as the second term in equation (6.6).

To calculate background coincidence levels C_{bkgd}, we instead take the average of all data points that satisfy the condition $y > D - C/4$, which seems to do an excellent job of accounting for the "humps" based on our observations. Nonetheless, we note that this value is simply an estimate, and as such should be considered an approximate value. The visibility of the dip, calculated by $C/(2C_{\text{bkgd}} + C)$, is thus also best considered approximate. Neither of these values are directly involved in the results, though they both serve as useful tools for evaluating the quality of a given trace and thus the fidelity of a given delay measurement.

Figure 6.12 shows the results of our measurements in the reflection geometry. Four sets of data are shown, each of which contains 176 data points representing individual

HOM traces. For each trace, coincidence was measured at 51 stage positions for 2 seconds, giving a total integration time per trace of 102 seconds. The black line represents the arithmetic mean of all four data points at each position. The reference voltage in reflection is $V_r = 0$ volts, and all delay measurements here are relative to the delay observed at that voltage.

The reflection delay $\partial\Phi_R/\partial\omega$ predicted by the simulation is shown in Figure 6.13 for liquid crystal voltages ranging from 1.5 V to 5 V, corresponding to rotation angles of $\phi = 69.3°$ to $\phi = 18.3°$ respectively. In this simulation the light is incident at 63.5° from N-SF11 glass ($n_{\mathrm{N-SF11}} = 1.76954$) on an 8-micron liquid crystal cell. The extraordinary and ordinary refractive indices used for the liquid crystal are $n_e = 1.633$ and $n_o = 1.54$. These indices are a little lower and higher, respectively, than the expected values for the E7 liquid crystal used in our experiment, and the cell thickness is a little thinner than the nominal value of 12 μm inferred from spectrophotometric measurements. However, these parameters primarily affect the spacing and magnitude of the Fabry-Perot fringes observed, and were chosen for better agreement with our experimental data. The cell contains a non-trivial amount of wedge, so it is not unreasonable to believe that the experiment took place at a different position on the cell that had a thickness closer to 8 microns or less.

The experimental data are in excellent agreement with the model predictions for $\tau_{\gamma,\mathrm{meas}}$, making it clear that the Δy contributions to the delay truly are suppressed in this type of measurement. The delay in the tunneling region (below approximately 2.2 V) appears to be identically zero within the experimental uncertainty, observed to be approximately $\pm$ 1 fs or less. Measurements below 1.5 V were consistent with these results as well, though they have been omitted from the plot for clarity. In the Fabry-Perot region, we observe the sharp dips at each resonance corresponding to interference from the Gaussian beam k-vector distribution as well as the slowly-increasing delay predicted between resonances. The sharp dips are not as pronounced as those shown in Figure 6.13 because the simulation only addresses the phase of

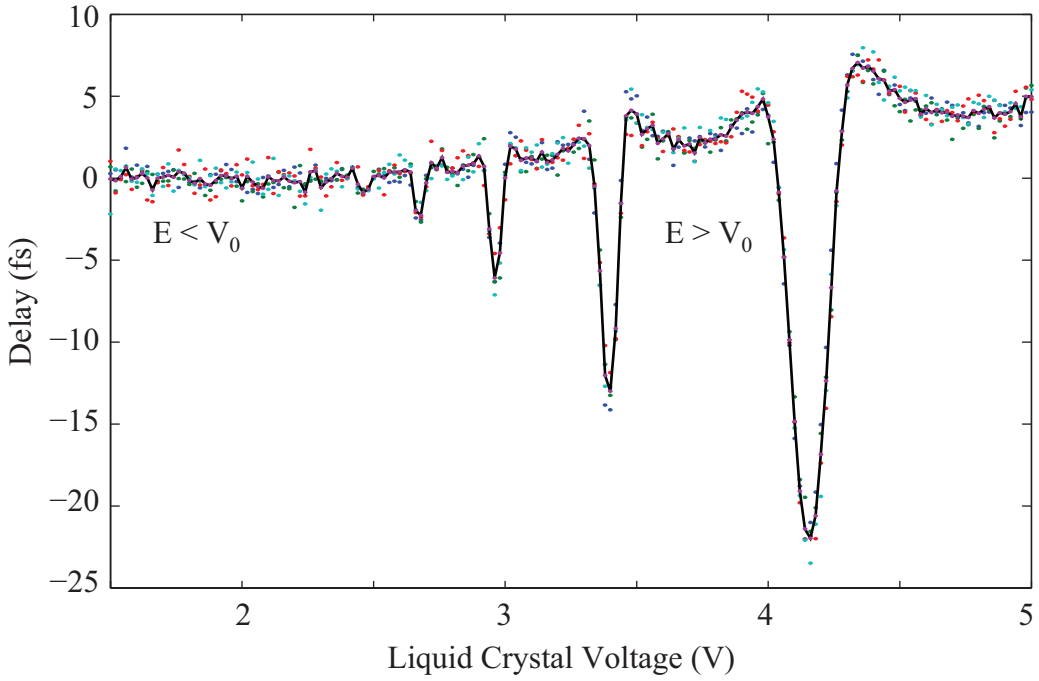

Figure 6.12: Reflection delay in FTIR from a double-prism barrier system. The black line is the arithmetic mean of four individual data sets, shown in colored dots. Each data point represents an individual Hong-Ou-Mandel trace with 102 seconds of integration time.

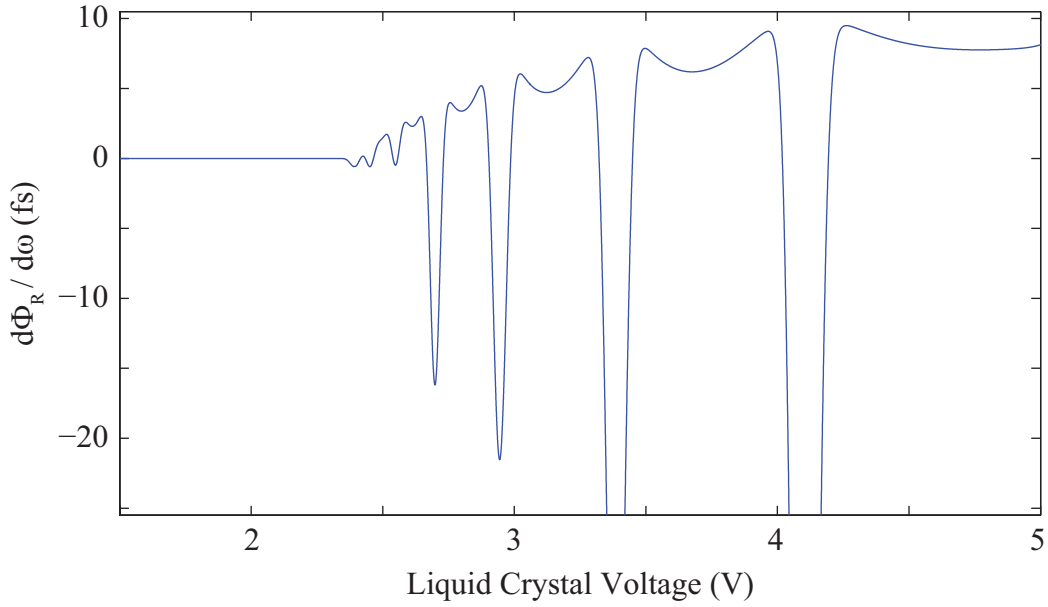

Figure 6.13: Predicted $\partial\Phi_R/\partial\omega$ based on the 4x4 matrix method described in the text for 727-nm light propagating through an 8-micron liquid crystal cell. The parameters used are $\theta = 63.5°$, $n_{\text{glass}} = 1.76954$, $n_e = 1.633$, and $n_o = 1.54$.

the reflection coefficient and weights each contribution equally. In experiment, the amplitude of the reflection coefficient is smaller at the Fabry-Perot resonances than when off-resonant, leading to an uneven weighting that reduces the magnitude of the dips in the measured delay.

Data taken in the transmission geometry is shown in Figure 6.14. Four data sets are shown, each of which has a successively larger integration time per stage position. The HOM trace time varies from 153 seconds for the first set to 4 hours and 15 minutes in the final set. The black line represents the weighted average according to integration time of all four data sets. In this case, the reference voltage is $V_r = 10$ volts, leading to an arbitrary delay offset as the reference voltage is located on one of the Fabry-Perot resonance lines. Despite this, the Fabry-Perot fringe effect is very clearly observed above the FTIR threshold. However at the lower voltages, where we're approaching the FTIR condition and entering the tunneling regime, the transmission drops significantly and we quickly become unable to fit the dip accurately. The effect of this transmission drop is immediately visible in the data as the voltage decreases from 3.5 volts to 2 volts. The variance or "spread" of the data points increases as voltage decreases, and below 2.2 volts the data becomes extremely unreliable. For voltages smaller than 2 volts, the data becomes bad enough that the fitting algorithm cannot give any useful information about the dynamics.

The transmission delay predicted by the numerical model is shown in Figure 6.15 for the same parameter values used in the reflection simulation. Fabry-Perot transmission resonances occur above the "critical voltage" where we transition from FTIR to allowed propagation as in the reflection plot, but in transmission there is no phase discontinuity at resonance and subsequently no anomalous delay features. These qualitative features of the model match the observed behavior, as expected.

However, there are some noticeable discrepancies between the data and the model that must be addressed. We do observe the expected Fabry-Perot effects all the way down to around 2.2 volts, but the peaks occur at different voltage values than the

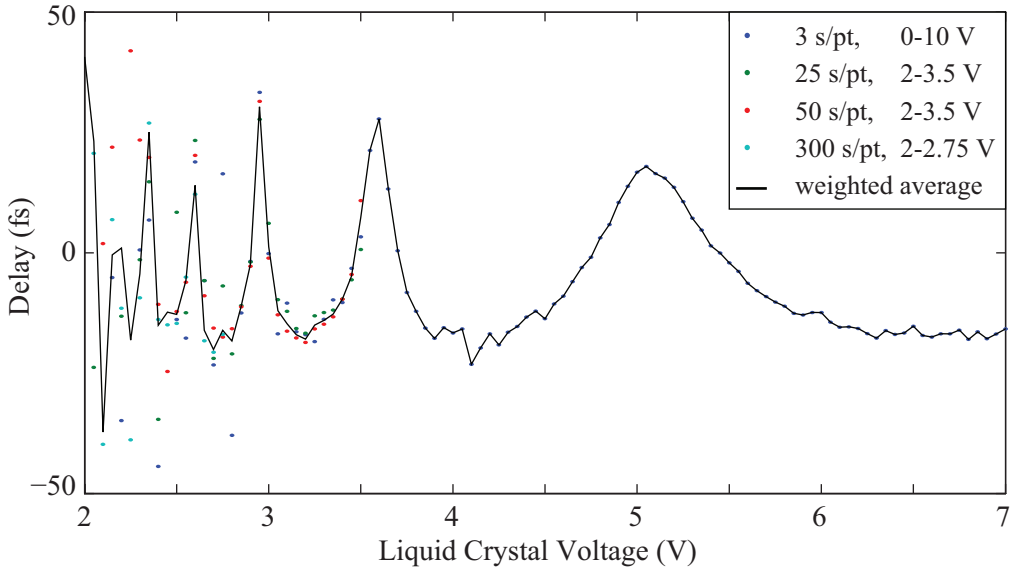

Figure 6.14: Transmission delay in FTIR through a double-prism barrier system. The black line is the weighted average according to integration time of the four individual data sets. Each data point represents a compiled Hong-Ou-Mandel trace from multiple interleaved measurements, as described in the text. The total integration time at each stage position (51 per trace) is shown in the legend.

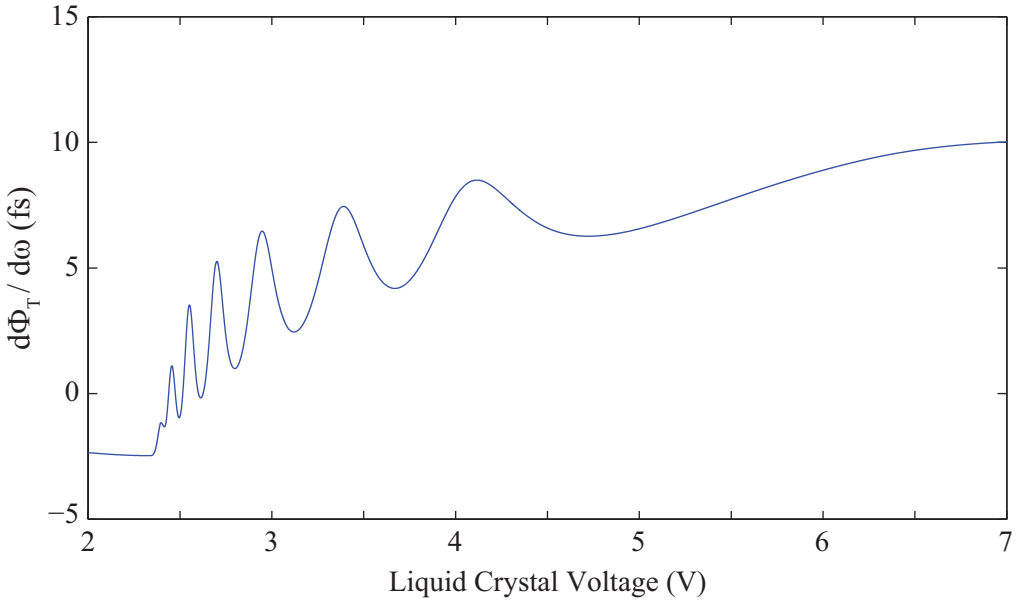

Figure 6.15: Predicted $\partial\Phi_T/\partial\omega$ from the 4x4 matrix method described in the text for the same parameters as shown in Figure 6.13.

model. This is a trivial effect that can be attributed to a change in the angle of incidence, which is difficult to avoid when reconfiguring the setup from the reflection geometry to the transmission geometry. In addition, we are likely interrogating a different section of the cell, which could lead to a different value of barrier thickness and/or cavity finesse. These two factors may also explain why the magnitude of the Fabry-Perot peaks is considerably larger than that predicted by the model. It is also possible that some sort of aberration is contributing, as it's not clear what a small amount of wedge would do to the Fabry-Perot features.

Below 2.2 volts, it is unclear whether we continue to see resonances, but the low transmission in this region suggests that FTIR has taken over for the majority of the beam, and we've transitioned into the tunneling regime. Without higher transmission or longer integration times, both of which are difficult if not impossible to obtain with our cell and experimental setup, we cannot obtain better resolution in this region. We hope to repeat the experiment with a much thinner cell in the future to increase transmission and obtain convincing tunneling delay data in transmission.

6.9 Conclusions

In this chapter, we have provided an extensive treatment of the double-prism FTIR system and identified a peculiarity of this system that as of yet has not been satisfactorily addressed in the literature. The suppression of the Δy contribution in the measurable portion of the tunneling delay in this system leads to differences between $\tau_{\gamma,\text{meas}}$ and τ_γ in excess of 50 fs or more, which are large enough to be easily verified by experimental measurements. The elimination of the transverse contributions caused by the Goos-Hänchen shift provides a deeper connection to the one-dimensional problem, as it isolates the longitudinal contributions that are characteristic of tunneling. By explicitly working out these relationships, we have identified a second possible mapping between the two problems.

We have also presented the first predictions for the observable delay in the double-prism FTIR system that properly incorporate the expected glass propagation time offset σ. Our predictions are based on a simulation using an existing 4x4 matrix technique that properly accounts for the anisotropy of the liquid crystal region acting as our photonic barrier. Scrutiny of these simulation results confirms that they conform to the conceptual predictions of standard Hartman theory; namely that the total delay τ_γ is the same for transmitted and reflected photons, and that the differences between $\tau_{\gamma,\text{meas}}$ in transmission and reflection are due to liquid crystal anisotropy.

These predictions have been confirmed by performing the first single-photon time delay measurements in a double-prism FTIR structure. In addition, these are the first direct time measurements of FTIR tunneling delay at optical frequencies. Since the results were obtained with single photons, they should put to rest many of the "reshaping" arguments that have been suggested as alternative explanations of the Hartman effect. The reflection results strongly suggest that the cavity interpretation presented by Winful is correct, and follow-up measurements with a thinner cell to confirm the model's transmission predictions should further support this interpretation.

Chapter 7

Conclusions

In this thesis, we have investigated two systems in which pulse propagation appears to occur superluminally. However, we have shown that in neither case is it necessary to ascribe the result to acausal sources, nor is it fair to claim that either system demonstrates a violation of Einstein causality. The work presented here attributes the observed pulse advancement to energy exchange or storage effects that differ from what is seen in normal propagation through a dielectric material, but these effects still fall well within the limits of causality.

7.1 Recurring themes

During our investigation we have encountered several recurring themes that relate to the apparent superluminality of these effects.

The first is that the peak of a pulse or wave packet is not the physically meaningful entity that carries information. The misinterpretation of these effects as "superluminal" is predicated on the idea that the peak represents something significant, generally either the position of the particle or the information carried by the particle. However, the peak of the wave packet is an interference effect that is not bound by causality or relativity. In any situation where that peak appears to advance acausally, it is simply

being re-created by interference between the components of the wave packet that were already present at the new location in spacetime. Once we disabuse ourselves of this notion of the peak as a physical entity that propagates causally, the vast majority of misinterpretations lose their substance.

Another recurring theme is that the information encoded in a pulse is contained in a point of non-analyticity. In that sense, a Gaussian pulse of infinite temporal extent cannot contain any information. It exists in all space and at all times, and any particular section contains all of the frequency components necessary to recreate the entire temporal pulse profile. The only way for such a pulse to contain information is through its presence or absence, which is impossible given the infinite nature of such a pulse (and suitably sensitive detectors).

Transmission of information occurs when the transmitter makes a decision and sends that decision to the receiver. That implies a time at which the pulse is "turned on" and emitted - a point of non-analyticity in the pulse profile. Generally, this turn-on point occurs far in advance of the pulse peak and at very low intensity, often below the detection threshold. However, experiments have shown that when that decision point is encoded on the higher-intensity portions of the pulse, often as a discontinuity in the pulse envelope, the discontinuity propagates causally at c [2, 30, 73].

From this we see why information propagation is effectively "capped" at c. We cannot observe energy prior to the causal arrival of the turn-on point. The peak of a truncated Gaussian pulse may arrive earlier or later than expected based on dispersive or energy-transfer effects, but it can never arrive earlier than the point of truncation, which is what would be necessary to violate special relativity.

Finally, we have seen that both of these propagation effects involve the distinction between energy in the propagating optical field and energy in a stored state, either an atomic excitation or an optical standing wave. In experiments, we usually only directly observe the propagating field energy. However, that field energy is intimately connected to the stored energy in the medium or cavity, which we generally do not

directly observe. The interaction and exchange of energy between propagating and stored states is critical to understanding these effects, because without that exchange the propagation ceases to appear superluminal.

In fact, one could extend this statement to cover most, if not all, propagation effects. Propagation through vacuum or free-space is merely a special case where no energy is exchanged with the system. Any other case of slow- or fast-light propagation can be explained as a time-dependent energy exchange between optical pulse and a material system or cavity. As a simple example, the delay observed on-resonance in a Fabry-Perot interferometer is due to the large energy build-up in the cavity standing wave. The cavity lifetime is very long when tuned to resonance, leading to a large delay. The opposite effect occurs when tuned between resonances, as destructive interference reduces the stored energy in the Fabry-Perot and decreases the cavity lifetime accordingly.

7.2 Specific findings

In conclusion, we have presented theoretical and experimental evidence that these two apparently superluminal propagation effects can be described by causal, luminal energy transfer between a propagating wave and a storage medium. We will now briefly summarize the specific findings of each section of the thesis.

In chapter 3, we investigated negative group velocities in an erbium-doped optical fiber (EDOF) system. We presented a rate-equation model that predicted pulse advancement and the presence of a backward-propagating peak within the EDOF, linking the incident and transmitted peaks at the entrance and exit faces of the fiber.

Experimentally, we have made the first experimental measurements of the pulse propagation dynamics in this system. Our measurements have confirmed that the energy velocity in such a system is always positive and thus not equal to the group velocity. Since the energy flow is in the forward direction, the effect cannot be the

result of interference between forward- and backward-propagating waves. Instead, it is due to a time-dependent interaction between a single, forward-propagating wave and the stored energy in the atomic excitation. We have also made the first direct observation of the backward-propagating peak predicted by the theoretical models.

These results are consistent with a time-dependent saturation of the erbium gain material that results in a strong energy exchange between material and optical field. The EDOF selectively amplifies the leading edge of the pulse, creating the advanced output peak as well as the backward-propagating peak within the fiber. This time-dependent energy exchange is responsible for the observed pulse advancement and the appearance of superluminal propagation.

In chapter 4 we reviewed the one-dimensional finite barrier problem and presented a cavity interpretation of tunneling based on the dwell time and self-interference delay. The standing wave in the cavity takes multiple round-trip propagation times to reach equilibrium, after which the cavity adiabatically responds to a change in input conditions on the time scale of the cavity lifetime. In this interpretation the observed delay is simply a cavity lifetime rather than a propagation delay, eliminating any conflict with the principle of Einstein causality. The misinterpretation of this cavity lifetime as a propagation delay, and insistence that the observed delay represents a violation of Einstein causality [4–7], was one of the largest motivating factors behind our work.

In chapter 5 we investigated a two-dimensional electromagnetic analog to the one-dimensional finite barrier problem in the form of frustrated total internal reflection (FTIR). We presented work performed by Steinberg and Chiao [82] in which they developed a partial mapping of the two-dimensional FTIR problem onto the one-dimensional finite barrier problem. By decomposing the group delay into dwell time and self-interference components, we were able to successfully complete the mapping and find an expression valid for arbitrary particle energy and barrier height (or, in the language of FTIR, index contrast and incidence angle). This decomposition divides

the process into a cavity-like delay in the tunneling region and a standing-wave effect in the region of incidence, and provides deeper insight into the energy dynamics of the tunneling process.

Finally, in chapter 6 we discussed the influence of input/output geometry on the measurable delay from a double-prism FTIR system. We have shown that in a measurement of this type the effects of the Goos-Hänchen shift are identically suppressed, a detail that has so far been overlooked in the literature. We have also presented the first accurate theoretical predictions of the delay expected from such a structure, including geometric effects. These predictions were compared to single-photon time delay measurements taken in reflection from a liquid-crystal-filled double-prism structure, and we observed good agreement between theory and experiment. This marks the first direct time measurement of reflection tunneling delays at optical frequencies. These results suggest that the cavity interpretation of tunneling delay is correct and that it is not appropriate to infer a propagation velocity from such a delay.

Bibliography

[1] J. Garrison, "Superluminal signals: causal loop paradoxes revisited," Physics Letters A **245**, 19–25 (1998).

[2] S. Zhang, J. Chen, C. Liu, M. Loy, G. Wong, and S. Du, "Optical Precursor of a Single Photon," Physical Review Letters **106**, 243602 (2011).

[3] A. Haibel and G. Nimtz, "Universal Relationship of Time and Frequency in Photonic Tunnelling," Annalen der Physik **10**, 707–712 (2001).

[4] G. Nimtz and A. A. Stahlhofen, "Macroscopic violation of special relativity," `arXiv:0708.0681v1 [quant-ph]` (2007).

[5] G. Nimtz and A. A. Stahlhofen, "How quantum tunneling broke the light speed barrier: Photonic Tunneling proceeds faster than Light," Phys. Chem. News **39**, 1–4 (2008).

[6] G. Nimtz, "Tunneling Violates Special Relativity," `arXiv:1003.3944v1 [quant-ph]` (2010).

[7] G. Nimtz, "Tunneling Confronts Special Relativity," Foundations of Physics **41**, 1193–1199 (2011).

[8] J. D. Jackson, *Classical Electrodynamics* (John Wiley & Sons, New York, 1999), 3rd ed.

[9] A. Kasapi, M. Jain, G. Yin, and S. Harris, "Electromagnetically Induced Transparency: Propagation Dynamics," Physical Review Letters **74**, 2447–2450 (1995).

[10] M. Kash, V. Sautenkov, A. Zibrov, L. Hollberg, G. Welch, M. Lukin, Y. Rostovtsev, E. Fry, and M. Scully, "Ultraslow Group Velocity and Enhanced Nonlinear Optical Effects in a Coherently Driven Hot Atomic Gas," Physical Review Letters **82**, 5229–5232 (1999).

[11] D. Budker, D. Kimball, S. Rochester, and V. Yashchuk, "Nonlinear Magnetooptics and Reduced Group Velocity of Light in Atomic Vapor with Slow Ground State Relaxation," Physical Review Letters **83**, 1767–1770 (1999).

[12] L. V. Hau, S. E. Harris, Z. Dutton, and C. H. Behroozi, "Light speed reduction to 17 metres per second in an ultracold atomic gas," Nature **397**, 594–598 (1999).

[13] C. Liu, Z. Dutton, C. H. Behroozi, and L. V. Hau, "Observation of coherent optical information storage in an atomic medium using halted light pulses." Nature **409**, 490–493 (2001).

[14] M. Bigelow, N. Lepeshkin, and R. W. Boyd, "Observation of Ultraslow Light Propagation in a Ruby Crystal at Room Temperature," Physical Review Letters **90**, 7–10 (2003).

[15] M. S. Bigelow, N. N. Lepeshkin, and R. W. Boyd, "Superluminal and slow light propagation in a room-temperature solid." Science (New York, N.Y.) **301**, 200–202 (2003).

[16] H. Gersen, T. Karle, R. Engelen, W. Bogaerts, J. Korterik, N. van Hulst, T. Krauss, and L. Kuipers, "Real-Space Observation of Ultraslow Light in Photonic Crystal Waveguides," Physical Review Letters **94**, 3–6 (2005).

[17] Y. A. Vlasov, M. O'Boyle, H. F. Hamann, and S. J. McNab, "Active control of slow light on a chip with photonic crystal waveguides." Nature **438**, 65–69 (2005).

[18] A. Yariv, Y. Xu, R. K. Lee, and A. Scherer, "Coupled-resonator optical waveguide: a proposal and analysis," Optics Letters **24**, 711 (1999).

[19] J. K. S. Poon, J. Scheuer, Y. Xu, and A. Yariv, "Designing coupled-resonator optical waveguide delay lines," Journal of the Optical Society of America B **21**, 1665 (2004).

[20] F. Xia, L. Sekaric, and Y. Vlasov, "Ultracompact optical buffers on a silicon chip," Nature Photonics **1**, 65–71 (2007).

[21] L. Brillouin, *Wave Propagation and Group Velocity* (Academic Press, New York, 1960).

[22] C. Garrett and D. McCumber, "Propagation of a Gaussian Light Pulse through an Anomalous Dispersion Medium," Physical Review A **1**, 305–313 (1970).

[23] R. Chiao, "Superluminal (but causal) propagation of wave packets in transparent media with inverted atomic populations," Physical Review A **48**, R34–R37 (1993).

[24] M. Ware, S. Glasgow, and J. Peatross, "Energy transport in linear dielectrics." Optics express **9**, 519–532 (2001).

[25] B. Macke and B. Segard, "Propagation of light-pulses at a negative group-velocity," The European Physical Journal D - Atomic, Molecular and Optical Physics **23**, 125–141 (2003).

[26] S. Chu and S. Wong, "Linear Pulse Propagation in an Absorbing Medium," Physical Review Letters **48**, 738–741 (1982).

[27] B. Segard and B. Macke, "Observation of negative velocity pulse propagation," Physics Letters A **109**, 213–216 (1985).

[28] L. Wang, A. Kuzmich, and A. Dogariu, "Gain-assisted superluminal light propagation," Nature **406**, 277–279 (2000).

[29] M. Talukder, Y. Amagishi, and M. Tomita, "Superluminal to Subluminal Transition in the Pulse Propagation in a Resonantly Absorbing Medium," Physical Review Letters **86**, 3546–3549 (2001).

[30] M. D. Stenner, D. J. Gauthier, and M. A. Neifeld, "The speed of information in a 'fast-light' optical medium." Nature **425**, 695–698 (2003).

[31] A. Schweinsberg, N. N. Lepeshkin, M. S. Bigelow, R. W. Boyd, and S. Jarabo, "Observation of superluminal and slow light propagation in erbium-doped optical fiber," Europhysics Letters (EPL) **73**, 218–224 (2006).

[32] R. d. L. Kronig, "On the theory of dispersion of x-rays," Journal of the Optical Society of America **12**, 547 (1926).

[33] H. Kramers, "La diffusion de la lumiere par les atomes," Atti Cong. Intern. Fisica, (Transactions of Volta Centenary Congress) Como **2**, 545–557 (1927).

[34] R. W. Boyd, *Nonlinear Optics* (Academic Press, San Diego, 2003), 2nd ed.

[35] Y. Okawachi, M. Bigelow, J. Sharping, Z. Zhu, A. Schweinsberg, D. Gauthier, R. Boyd, and A. Gaeta, "Tunable All-Optical Delays via Brillouin Slow Light in an Optical Fiber," Physical Review Letters **94**, 153902 (2005).

[36] J. E. Sharping, Y. Okawachi, and A. L. Gaeta, "Wide bandwidth slow light using a Raman fiber amplifier," Optics Express **13**, 6092 (2005).

[37] Y. Okawachi, M. Foster, J. Sharping, A. Gaeta, Q. Xu, and M. Lipson, "All-optical slow-light on a photonic chip," Optics Express **14**, 2317 (2006).

[38] A. Dawes, D. Gauthier, and A. Willner, "12-GHz-Bandwidth SBS Slow Light in Optical Fibers," in "Optical Fiber Communication Conference/National Fiber Optic Engineers Conference, 2006," (IEEE, 2006), pp. 1–3.

[39] H. Su and S. L. Chuang, "Room-temperature slow light with semiconductor quantum-dot devices," Optics Letters **31**, 271 (2006).

[40] Z. Shi, R. Pant, Z. Zhu, M. D. Stenner, M. A. Neifeld, D. J. Gauthier, and R. W. Boyd, "Design of a tunable time-delay element using multiple gain lines for increased fractional delay with high data fidelity," Optics Letters **32**, 1986 (2007).

[41] R. Camacho, M. Pack, J. Howell, A. Schweinsberg, and R. Boyd, "Wide-Bandwidth, Tunable, Multiple-Pulse-Width Optical Delays Using Slow Light in Cesium Vapor," Physical Review Letters **98**, 153601 (2007).

[42] R. W. Boyd and D. J. Gauthier, "'Slow' and 'fast' light," in "Progress in Optics," , E. Wolf, ed. (Elsevier Science, Amsterdam, 2002), chap. 6, pp. 497–530.

[43] S. Ghosh, J. Sharping, D. Ouzounov, and A. L. Gaeta, "Resonant Optical Interactions with Molecules Confined in Photonic Band-Gap Fibers," Physical Review Letters **94**, 093902 (2005).

[44] S. Ghosh, A. Bhagwat, C. Renshaw, S. Goh, A. L. Gaeta, and B. Kirby, "Low-Light-Level Optical Interactions with Rubidium Vapor in a Photonic Band-Gap Fiber," Physical Review Letters **97**, 023603 (2006).

[45] V. Boyer, C. McCormick, E. Arimondo, and P. Lett, "Ultraslow Propagation of Matched Pulses by Four-Wave Mixing in an Atomic Vapor," Physical Review Letters **99**, 143601 (2007).

[46] P.-C. Ku, F. Sedgwick, C. J. Chang-Hasnain, P. Palinginis, T. Li, H. Wang, S.-W. Chang, and S.-L. Chuang, "Slow light in semiconductor quantum wells," Optics Letters **29**, 2291 (2004).

[47] P. Palinginis, F. Sedgwick, S. Crankshaw, M. Moewe, and C. J. Chang-Hasnain, "Room temperature slow light in a quantum-well waveguide via coherent population oscillation," Optics Express **13**, 9909 (2005).

[48] P. Palinginis, S. Crankshaw, F. Sedgwick, E.-T. Kim, M. Moewe, C. J. Chang-Hasnain, H. Wang, and S.-L. Chuang, "Ultraslow light (<200 m/s) propagation in a semiconductor nanostructure," Applied Physics Letters **87**, 171102 (2005).

[49] J. Mørk, R. Kjær, M. van der Poel, and K. Yvind, "Slow light in a semiconductor waveguide at gigahertz frequencies," Optics Express **13**, 8136 (2005).

[50] M. L. Povinelli, S. G. Johnson, and J. D. Joannopoulos, "Slow-light, band-edge waveguides for tunable time delays," Optics Express **13**, 7145 (2005).

[51] S. P. Anderson, A. R. Shroff, and P. M. Fauchet, "Slow Light with Photonic Crystals for On-Chip Optical Interconnects," Advances in Optical Technologies **2008**, 293531 (2008).

[52] G. Lenz, B. Eggleton, C. Madsen, and R. Slusher, "Optical delay lines based on optical filters," IEEE Journal of Quantum Electronics **37**, 525–532 (2001).

[53] J. T. Mok, C. M. de Sterke, I. C. M. Littler, and B. J. Eggleton, "Dispersionless slow light using gap solitons," Nature Physics **2**, 775–780 (2006).

[54] J. T. Mok, M. Ibsen, C. M. de Sterke, and B. J. Eggleton, "Slow Light Generation Using Fibre Bragg Gratings," in "Optical Fiber Communication/National Fiber Optic Engineers Conference, 2008," (IEEE, 2008), pp. 1–3.

[55] J. E. Heebner, R. Boyd, and Q.-H. Park, "Slow light, induced dispersion, enhanced nonlinearity, and optical solitons in a resonator-array waveguide," Physical Review E **65**, 036619 (2002).

[56] D. Smith, H. Chang, K. Fuller, A. Rosenberger, and R. Boyd, "Coupled-resonator-induced transparency," Physical Review A **69**, 063804 (2004).

[57] J. Heebner, a. Schweinsberg, R. W. Boyd, and D. Jackson, "Optical transmission characteristics of fiber ring resonators," IEEE Journal of Quantum Electronics **40**, 726–730 (2004).

[58] K. Totsuka, N. Kobayashi, and M. Tomita, "Slow Light in Coupled-Resonator-Induced Transparency," Physical Review Letters **98**, 213904 (2007).

[59] F. Shinobu, N. Ishikura, Y. Arita, T. Tamanuki, and T. Baba, "Continuously tunable slow-light device consisting of heater-controlled silicon microring array," Optics Express **19**, 13557 (2011).

[60] H. Altug and J. Vuckovic, "Experimental demonstration of the slow group velocity of light in two-dimensional coupled photonic crystal microcavity arrays," Applied Physics Letters **86**, 111102 (2005).

[61] J. Scheuer, G. T. Paloczi, J. K. S. Poon, and A. Yariv, "Coupled Resonator Optical Waveguides: Toward the Slowing and Storage of Light," Optics and Photonics News **16**, 36 (2005).

[62] M. L. Cooper, G. Gupta, M. A. Schneider, W. M. J. Green, S. Assefa, F. Xia, D. K. Gifford, and S. Mookherjea, "Waveguide dispersion effects in silicon-on-insulator coupled-resonator optical waveguides," Optics Letters **35**, 3030 (2010).

[63] J. Kim, S. L. Chuang, P. C. Ku, and C. J. Chang-Hasnain, "Slow light using semiconductor quantum dots," Journal of Physics: Condensed Matter **16**, S3727–S3735 (2004).

[64] V. S. C. Managa Rao and S. Hughes, "Single quantum dots for slow and fast light in a planar photonic crystal," Optics Letters **32**, 304 (2007).

[65] R. W. Boyd, D. Gauthier, A. Gaeta, and A. Willner, "Maximum time delay achievable on propagation through a slow-light medium," Physical Review A **71**, 023801 (2005).

[66] H. Shin, A. Schweinsberg, G. Gehring, K. Schwertz, H. J. Chang, R. W. Boyd, Q.-h. Park, and D. J. Gauthier, "Reducing pulse distortion in fast-light pulse propagation through an erbium-doped fiber amplifier," Optics Letters **32**, 906 (2007).

[67] H. Shin, A. Schweinsberg, and R. W. Boyd, "Reducing pulse distortion in fast-light pulse propagation through an erbium-doped fiber amplifier using a mutually incoherent background field," Optics Communications **282**, 2085–2087 (2009).

[68] B. Zhang, L. Yan, I. Fazal, L. Zhang, A. E. Willner, Z. Zhu, and D. J. Gauthier, "Slow light on Gbit/s differential-phase-shift-keying signals." Optics express **15**, 1878 (2007).

[69] L. Zhang, T. Luo, C. Yu, W. Zhang, and A. E. Willner, "Pattern Dependence of Data Distortion in Slow-Light Elements," Journal of Lightwave Technology **25**, 1754–1760 (2007).

[70] L. Brillouin, "Über die Fortpflanzung des Lichtes in dispergierenden Medien," Annalen der Physik **349**, 203–240 (1914).

[71] A. Sommerfeld, "Über die Fortpflanzung des Lichtes in dispergierenden Medien," Annalen der Physik **349**, 177–202 (1914).

[72] R. L. Smith, "The Velocities of Light," American Journal of Physics **38**, 978 (1970).

[73] M. Stenner, D. Gauthier, and M. Neifeld, "Fast Causal Information Transmission in a Medium With a Slow Group Velocity," Physical Review Letters **94**, 053902 (2005).

[74] N. G. Basov, R. V. Ambartsumyan, V. S. Zuev, P. G. Kryukov, and V. S. Letokhov, "Nonlinear Amplification of Light Pulses," Sov. Phys. JETP **23**, 16 (1966).

[75] G. Sherman and K. Oughstun, "Description of Pulse Dynamics in Lorentz Media in Terms of the Energy Velocity and Attenuation of Time-Harmonic Waves," Physical Review Letters **47**, 1451–1454 (1981).

[76] M. Ware, S. Glasgow, and J. Peatross, "Role of group velocity in tracking field energy in linear dielectrics." Optics express **9**, 506–518 (2001).

[77] G. Agarwal and T. Dey, "Sub- and Superluminal Propagation of Intense Pulses in Media with Saturated and Reverse Absorption," Physical Review Letters **92**, 203901 (2004).

[78] S. Chu and S. Wong, "Chu and Wong Respond," Physical Review Letters **49**, 1293–1293 (1982).

[79] L. Hillman, "Observation of a spectral hole due to population oscillations in a homogeneously broadened optical absorption line," Optics Communications **45**, 416–419 (1983).

[80] S. Novak and A. Moesle, "Analytic Model for Gain Modulation in EDFAs," Journal of Lightwave Technology **20**, 975–985 (2002).

[81] G. Diener, "Energy balance and energy transport velocity in dispersive media," Annalen der Physik **7**, 639–644 (1998).

[82] A. M. Steinberg and R. Chiao, "Tunneling delay times in one and two dimensions," Physical Review A **49**, 3283–3295 (1994).

[83] H. Winful, "Tunneling time, the Hartman effect, and superluminality: A proposed resolution of an old paradox," Physics Reports **436**, 1–69 (2006).

[84] E. Hauge and J. Støvneng, "Tunneling times: a critical review," Reviews of Modern Physics **61**, 917–936 (1989).

[85] R. Landauer and T. Martin, "Barrier interaction time in tunneling," Reviews of Modern Physics **66**, 217–228 (1994).

[86] H. Winful, "Delay Time and the Hartman Effect in Quantum Tunneling," Physical Review Letters **91**, 260401 (2003).

[87] H. G. Winful, "The meaning of group delay in barrier tunnelling: a reexamination of superluminal group velocities," New Journal of Physics **8**, 101 (2006).

[88] G. Torrese, H. P. Schriemer, and M. Cada, "Pulse propagation in finite linear one-dimensional periodic structures," Proceedings of SPIE **5577**, 568–578 (2004).

[89] M. Büttiker and R. Landauer, "Traversal Time for Tunneling," Physical Review Letters **49**, 1739–1742 (1982).

[90] S. Collins, D. Lowe, and J. R. Barker, "The quantum mechanical tunnelling time problem-revisited," Journal of Physics C: Solid State Physics **20**, 6213–6232 (1987).

[91] A. M. Steinberg, "How much time does a tunneling particle spend in the barrier region?" Physical review letters **74**, 2405–2409 (1995).

[92] J. Falck and E. Hauge, "Larmor clock reexamined," Physical Review B **38**, 3287–3297 (1988).

[93] M. Azbel, "Eigenstates and properties of random systems in one dimension at zero temperature," Physical Review B **28**, 4106–4125 (1983).

[94] C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Quantum Mechanics* (Wiley-Interscience, New York, 1977).

[95] P. Balcou and L. Dutriaux, "Dual Optical Tunneling Times in Frustrated Total Internal Reflection," Physical Review Letters **78**, 851–854 (1997).

[96] T. E. Hartman, "Tunneling of a Wave Packet," Journal of Applied Physics **33**, 3427 (1962).

[97] V. S. Olkhovsky and E. Recami, "Recent developments in the time analysis of tunneling processes," Physics Reports **214**, 339–356 (1992).

[98] H. Winful, M. Ngom, and N. Litchinitser, "Relation between quantum tunneling times for relativistic particles," Physical Review A **70**, 052112 (2004).

[99] A. M. Steinberg, P. Kwiat, and R. Chiao, "Measurement of the single-photon tunneling time," Physical Review Letters **71**, 708–711 (1993).

[100] A. M. Steinberg and R. Chiao, "Subfemtosecond determination of transmission delay times for a dielectric mirror (photonic band gap) as a function of the angle of incidence," Physical Review A **51**, 3525–3528 (1995).

[101] C. Spielmann, R. Szipöcs, A. Stingl, and F. Krausz, "Tunneling of Optical Pulses through Photonic Band Gaps," Physical Review Letters **73**, 2308–2311 (1994).

[102] S. Longhi, M. Marano, P. Laporta, and M. Belmonte, "Superluminal optical pulse propagation at 1.5 μm in periodic fiber Bragg gratings," Physical Review E **64**, 055602 (2001).

[103] M. Mojahedi, E. Schamiloglu, F. Hegeler, and K. Malloy, "Time-domain detection of superluminal group velocity for single microwave pulses," Physical Review E **62**, 5758–5766 (2000).

[104] M. Mojahedi, K. Malloy, G. Eleftheriades, J. Woodley, and R. Chiao, "Abnormal wave propagation in passive media," IEEE Journal of Selected Topics in Quantum Electronics **9**, 30–39 (2003).

[105] A. Hache and L. Poirier, "Long-range superluminal pulse propagation in a coaxial photonic crystal," Applied Physics Letters **80**, 518 (2002).

[106] M. Reiten, D. Grischkowsky, and R. Cheville, "Optical tunneling of single-cycle terahertz bandwidth pulses," Physical Review E **64**, 036604 (2001).

[107] S. Yang, J. Page, Z. Liu, M. Cowan, C. Chan, and P. Sheng, "Ultrasound Tunneling through 3D Phononic Crystals," Physical Review Letters **88**, 104301 (2002).

[108] W. M. Robertson, J. Ash, and J. M. McGaugh, "Breaking the sound barrier: Tunneling of acoustic waves through the forbidden transmission region of a one-dimensional acoustic band gap array," American Journal of Physics **70**, 689 (2002).

[109] F. Smith, "Lifetime Matrix in Collision Theory," Physical Review **118**, 349–356 (1960).

[110] M. Büttiker, "Larmor precession and the traversal time for tunneling," Physical Review B **27**, 6178–6188 (1983).

[111] C. Leavens and G. Aers, "Dwell time and phase times for transmission and reflection," Physical Review B **39**, 1202–1206 (1989).

[112] E. Hauge, J. Falck, and T. Fjeldly, "Transmission and reflection times for scattering of wave packets off tunneling barriers," Physical Review B **36**, 4203–4214 (1987).

[113] T. Martin and R. Landauer, "Time delay of evanescent electromagnetic waves and the analogy to particle tunneling," Physical Review A **45**, 2611–2617 (1992).

[114] H. Winful, "Group delay, stored energy, and the tunneling of evanescent electromagnetic waves," Physical Review E **68**, 016615 (2003).

[115] C. A. Balanis, *Advanced Engineering Electromagnetics* (John Wiley & Sons, 1989).

[116] A. Ghatak and S. Banerjee, "Temporal delay of a pulse undergoing frustrated total internal reflection," Applied Optics **28**, 1960 (1989).

[117] B. Lee and W. Lee, "TM-polarized photon tunneling phase time in a frustrated-total-internal-reflection structure," Journal of the Optical Society of America B **14**, 777 (1997).

[118] R. Chiao, P. Kwiat, and A. M. Steinberg, "Analogies between electron and photon tunneling: A proposed experiment to measure photon tunneling times," Physica B: Condensed Matter **175**, 257–262 (1991).

[119] A. B. Shvartsburg, "Tunneling of electromagnetic waves: paradoxes and prospects," Physics-Uspekhi **50**, 37–51 (2007).

[120] J. C. Martinez and E. Polatdemir, "Origin of the Hartman effect," Physics Letters A **351**, 31–36 (2006).

[121] J. C. Martinez, E. Polatdemir, and M. Taufik, "Interference and dispersion effects on tunneling time," The European Physical Journal D **41**, 49–54 (2006).

[122] D. J. Papoular, P. Clade, S. V. Polyakov, C. F. McCormick, A. L. Migdall, and P. D. Lett, "Measuring optical tunneling times using a Hong-Ou-Mandel interferometer," Optics Express **16**, 16005 (2008).

[123] M. Deutsch and J. Golub, "Optical Larmor clock: Measurement of the photonic tunneling time," Physical Review A **53**, 434–439 (1996).

[124] E. Pollak and W. Miller, "New Physical Interpretation for Time in Scattering Theory," Physical Review Letters **53**, 115–118 (1984).

[125] Schott North America, "Data and Tools for Optical Designers," `http://www.us.schott.com/advanced_optics/english/our_products/materials/data_tools/index.html` (2011).

[126] J. M. Geary, J. W. Goodby, A. R. Kmetz, and J. S. Patel, "The mechanism of polymer alignment of liquid-crystal materials," Journal of Applied Physics **62**, 4100 (1987).

[127] I.-C. Khoo, *Liquid Crystals* (Wiley-Interscience, Hoboken, NJ, 2007), 2nd ed.

[128] J. Li, C.-H. Wen, S. Gauza, R. Lu, and S.-T. Wu, "Refractive Indices of Liquid Crystals for Display Applications," Journal of Display Technology **1**, 51–61 (2005).

[129] S.-T. Wu, U. Efron, and L. D. Hess, "Birefringence measurements of liquid crystals," Applied Optics **23**, 3911 (1984).

[130] F. Brehat and B. Wyncke, "Reflectivity, transmissivity and optical constants of anisotropic absorbing crystals," Journal of Physics D: Applied Physics **24**, 2055–2066 (1991).

[131] S. Teitler and B. W. Henvis, "Refraction in Stratified, Anisotropic Media," Journal of the Optical Society of America **60**, 830 (1970).

[132] D. Berreman and T. Scheffer, "Bragg Reflection of Light from Single-Domain Cholesteric Liquid-Crystal Films," Physical Review Letters **25**, 577–581 (1970).

[133] D. W. Berreman, "Optics in Stratified and Anisotropic Media: 4 x 4-Matrix Formulation," Journal of the Optical Society of America **62**, 502 (1972).

[134] J. Lekner, "Reflection and refraction by uniaxial crystals," Journal of Physics: Condensed Matter **3**, 6121–6133 (1991).

[135] G. D. Landry and T. A. Maldonado, "Gaussian beam transmission and reflection from a general anisotropic multilayer structure," Applied Optics **35**, 5870 (1996).

[136] B. J. J. Slagmolen, M. B. Gray, K. G. Baigent, and D. E. McClelland, "Phase-Sensitive Reflection Technique for Characterization of a Fabry-Perot Interferometer," Applied Optics **39**, 3638 (2000).

[137] Z. Wang, C. Wang, and Z. Zhang, "Goos-Hänchen shift of the uniaxially anisotropic left-handed material film with an arbitrary angle between the optical axis and the interface," Optics Communications **281**, 3019–3024 (2008).

[138] C. K. Hong, Z. Y. Ou, and L. Mandel, "Measurement of subpicosecond time intervals between two photons by interference," Physical Review Letters **59**, 2044–2046 (1987).

[139] Z. Y. Ou and L. Mandel, "Observation of Spatial Quantum Beating with Separated Photodetectors," Physical Review Letters **61**, 54–57 (1988).

[140] A. M. Steinberg, P. Kwiat, and R. Chiao, "Dispersion cancellation in a measurement of the single-photon propagation velocity in glass," Physical Review Letters **68**, 2421–2424 (1992).

[141] A. M. Steinberg, P. Kwiat, and R. Chiao, "Dispersion cancellation and high-resolution time measurements in a fourth-order optical interferometer," Physical Review A **45**, 6659–6665 (1992).

[142] A. K. Jha, M. OSullivan, K. W. C. Chan, and R. W. Boyd, "Temporal coherence and indistinguishability in two-photon interference effects," Physical Review A **77**, 021801 (2008).

[143] R. Ghosh, C. Hong, Z. Ou, and L. Mandel, "Interference of two photons in parametric down conversion," Physical Review A **34**, 3962–3968 (1986).

[144] Perkin Elmer, "Single Photon Counting Modules (SPCM)," `http://www.perkinelmer.com/Category/Category/cat1/IDSMI_TAXONOMY_DELETIONS/cat2/IND_SE_CAT_Single%20Photon%20Counting%20Modules%20SPCM_001/key/10613` (2011).

[145] D. Branning, S. Bhandari, and M. Beck, "Low-cost coincidence-counting electronics for undergraduate quantum optics," American Journal of Physics **77**, 667 (2009).

[146] Beck, M. and D. Branning, "Coincidence Counting Units (CCUs)," `http://people.whitman.edu/~beckmk/QM/circuit/circuit.html` (2011).