Tunneling delays in frustrated total internal reflection

George M. Gehring,[1] Andreas C. Liapis,[1] and Robert W. Boyd[1,2,3]
[1]*The Institute of Optics, University of Rochester, Rochester, New York 14627, USA*
[2]*Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627, USA*
[3]*Department of Physics and School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5*
(Received 6 February 2012; published 22 March 2012)

We demonstrate a direct relationship between tunneling delays in a one-dimensional potential barrier and the delays observed in frustrated total internal reflection. Expressions relating the group delays in these two cases have been formulated in the past [A. M. Steinberg and R. Y. Chiao, Phys. Rev. A **49**, 3283 (1994)], but only for certain limits of the normalized particle energy E/V_0. Here, we derive a single expression relating these two group delays that is valid for arbitrary particle energy by decomposing the group delay into dwell-time and self-interference components. This decomposition also explains why the predicted delay differs in the limiting cases observed by Steinberg and Chiao.

DOI: 10.1103/PhysRevA.85.032122 PACS number(s): 03.65.Xp, 41.20.Jb, 42.25.Gy

I. INTRODUCTION

The delay experienced by a particle tunneling through a barrier has been a topic of scrutiny for the better part of a century [1–4]. The traditional definition of the group delay τ_g acquired when tunneling through a barrier of length L is calculated by the method of stationary phase and is frequently called the phase time or Wigner time. The group delay can surprisingly become shorter than the "equal time" c/L, suggesting superluminal transit of the tunneled particle. Other definitions, including a "semiclassical time" and "dwell time," have been proposed to try and resolve this apparent paradox [5].

More recently, Winful has provided an interpretation of tunneling delays that emphasizes the cavity-like nature of the phenomenon [6]. He emphasizes the recognition of tunneling as a quasistatic process and the interpretation of the group delay as a cavity lifetime rather than a transit time. His analysis decomposes the group delay into two more fundamental components: the dwell time τ_d and the self-interference delay τ_i. The dwell time represents the average time spent in the barrier by a particle and is reminiscent of a collision time [5]. The self-interference delay represents additional delay caused by the interference of the incident and reflected parts of the particle wave function in front of the barrier and is proportional to reactive stored energy.

The group delay in the case of two-dimensional (2D) tunneling of electromagnetic waves in frustrated total internal reflection (FTIR), has been worked out by several authors for the simple case of a glass-air-glass geometry [7–9]. The similarities between the 2D tunneling and one-dimensional (1D) tunneling problems have also been noted [10]. Steinberg and Chiao performed the first one-to-one mapping of the two-dimensional FTIR problem onto the one-dimensional quantum-mechanical problem but found that the mapping could only be completed by approximation in two limits of the one-dimensional particle energy E and barrier potential V_0, namely, $E \ll V_0$ and $E \geqslant V_0$ [8]. Curiously, the mapping was qualitatively different for the two limits of particle energy.

In this paper we will demonstrate that, by utilizing the dwell-time and self-interference-delay decomposition of Winful, we can derive a more accurate mapping between the 2D and 1D tunneling problems. This mapping produces results consistent with those of Steinberg and Chiao in the appropriate limits but is no longer limited to those special cases. Rather, our expression provides a complete description of the process for any value of 1D particle energy E and illustrates the changes that occur between the limiting cases observed previously. We feel that our interpretation provides a deeper physical insight into the tunneling phenomenon.

We will begin by presenting an overview of prior work mapping 2D tunneling onto the 1D tunneling problem and reviewing the canonical definitions of the dwell time and self-interference delay. These definitions will then be used to perform a decomposition of the 2D tunneling delay for TE and TM boundary conditions and to demonstrate that such a decomposition leads to a one-to-one mapping between the 2D and 1D tunneling delays. Finally, we will discuss a practical issue that dictates which portions of the group delay can and cannot be measured in the double-prism FTIR geometry proposed in [8].

II. PREVIOUS WORK

The 1D tunneling problem is thoroughly covered in [11]. The system is shown in Fig. 1. A particle of energy E is incident from a region of zero potential energy (region I) upon a barrier of potential energy V_0 that extends from $x = 0$ to $x = L$ (region II). The potential energy returns to $V = 0$ for $x > L$ (region III). The stationary state wave functions in each region are found from the time-independent Schrödinger equation and have the following form:

$$\Psi_{\rm I}(x) = e^{ikx} + Re^{-ikx}, \tag{1a}$$

$$\Psi_{\rm II}(x) = Ce^{-\kappa x} + De^{\kappa x}, \tag{1b}$$

$$\Psi_{\rm III}(x) = Te^{ikx}. \tag{1c}$$

In these equations we have explicitly defined the particle wave vector k in regions I and III ($i\kappa$ in region II) in the usual

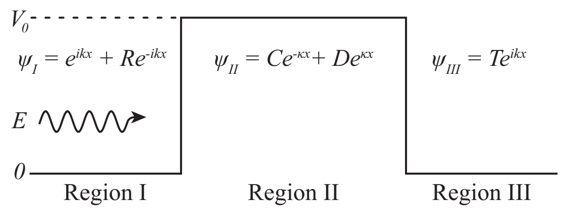

FIG. 1. Schematic diagram for one-dimensional barrier tunneling. A particle of energy E approaches a barrier of potential V_0. ψ is the wave function in each region.

fashion, with $E = \hbar^2 k^2/2m$ and $\kappa = \sqrt{2m(V_0 - E)/\hbar}$ for a particle of effective mass m. Finding the coefficients R, T, C, and D is a straightforward exercise in algebra. We will quote the results, continuing to use a notation similar to that given in [11,12] for consistency:

$$R = -\frac{i\Delta'}{g}\sinh\kappa L, \tag{2a}$$

$$T = \frac{e^{-ikL}}{g}, \tag{2b}$$

$$C = \left(1 - \frac{ik}{\kappa}\right)e^{\kappa L}/2g, \tag{2c}$$

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

where several parameters have been defined as follows to simplify the notation:

$$\Delta = \frac{1}{2}\left(\frac{\kappa}{k} - \frac{k}{\kappa}\right), \tag{3a}$$

$$\Delta' = \frac{1}{2}\left(\frac{\kappa}{k} + \frac{k}{\kappa}\right), \tag{3b}$$

$$g = \cosh\kappa L + i\Delta\sinh\kappa L. \tag{3c}$$

A spatially localized wave packet is constructed from a narrow band of stationary states. The group delay, which measures the time between the arrival of the wave-packet peak at $x = 0$ and its appearance at $x = L$, can then be calculated by applying the method of stationary phase. The group delay in transmission τ_{gt} is calculated by taking the energy derivative of the transmission phase shift,

$$\tau_{gt} = \hbar\frac{d\phi_0}{dE}, \tag{4}$$

where $\phi_0 = \phi_t + kL$ and $\phi_t = \arg(T)$. A similar expression can be found for the group delay in reflection,

$$\tau_{gr} = \hbar\frac{d\phi_r}{dE},$$

where $\phi_r = \arg(R)$. It can be shown that for a symmetric barrier, $\tau_{gt} = \tau_{gr}$ [13–15]. We will use τ_g to represent the group delay calculated with this Wigner phase-time method for a symmetric barrier.

The transmission phase ϕ_0 can be calculated from T as given in (2b), yielding

$$\phi_0 = -\arctan(\Delta\tanh\kappa L).$$

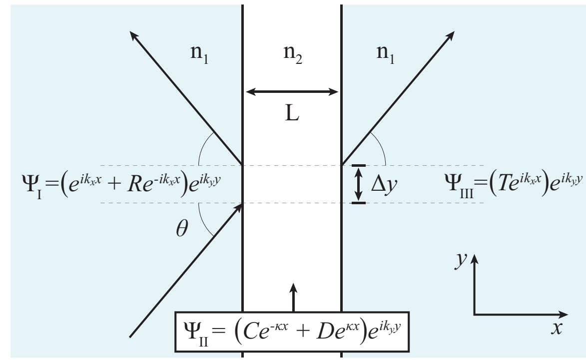

FIG. 2. (Color online) Schematic diagram for two-dimensional barrier tunneling.

Differentiating this phase with respect to energy and simplifying gives us an expression for τ_g in terms of k, κ, m, and L,

$$\tau_g = \frac{mL}{\hbar k}\frac{\cos^2\phi_0}{2} \times \left[\left(\frac{\kappa}{k} + \frac{k}{\kappa}\right)^2\frac{\tanh\kappa L}{\kappa L} + \left(1 - \frac{k^2}{\kappa^2}\right)\text{sech}^2\kappa L\right]. \tag{5}$$

To draw analogies between the 1D tunneling problem and FTIR, Steinberg and Chiao consider a system such as that shown in Fig. 2. A slab of material with index n_2 and width L is sandwiched between two regions of material of index n_1. In [8], Steinberg and Chiao consider a glass-air-glass system with $n_2 = 1$; however, it will be convenient for us to keep this medium arbitrary for later calculations. We will label regions I, II, and III from left to right as in the one-dimensional calculation, with region II 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 dashed lines in Fig. 2 represent the planes $y = 0$ and $y = \Delta y$.

For total internal reflection to occur, Snell's law requires that $n_1\sin\theta > n_2$, or equivalently 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 a lateral shift Δy.

The electric field of a TE-polarized (perpendicular to the plane of incidence) plane wave can be represented in regions I–III as

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

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. Substitution of this form into the Helmholtz equation demonstrates that the solutions of $\psi(x)$ in the 2D problem are identical to those given in Eq. (1) with $k \to k_x$. However, in the 2D problem k_x and κ are given in terms of the refractive indices n_1 and n_2 as well as the incidence angle θ:

$$k_x^2 = \frac{\omega^2}{c^2}n_1^2\cos^2\theta \quad \text{for} \quad \text{I, III}, \tag{6a}$$

$$\kappa^2 = \frac{\omega^2}{c^2}\left(n_1^2\sin^2\theta - n_2^2\right) \quad \text{for} \quad \text{II}. \tag{6b}$$

TABLE I. Coefficient equivalences.

Wave vector	1D	2D
k_x^2	$2mE/\hbar^2$	$\frac{n_1^2\omega^2}{c^2}\cos^2\theta$
κ^2	$2m(V_0-E)/\hbar^2$	$\frac{\omega^2}{c^2}(n_1^2\sin^2\theta - n_2^2)$

Continuity of E_z across the boundary forces $\psi(x)$ to be continuous, and with the assumption that the relative permeability $\mu_r = 1$, continuity of B_y forces $\partial\psi(x)/\partial x$ to be continuous. Thus, in this special case, the boundary conditions are identical to those of the Schrödinger equation for the 1D problem. As a result, the previous expressions for R, T, C, and D apply to the two-dimensional problem as well with the proper choice of k_x and κ. This equivalence between the 1D and 2D problems is summarized in Table I.

The transmitted phase is then $\Phi = \phi_t + k_x L + k_y y - \omega t$, with ϕ_t defined as before. In two dimensions, the stationary phase approximation dictates that the gradient of the phase in k space should vanish. This gradient can be taken with respect to the magnitude and direction of the k vector, giving two constraint equations that describe the evolution of the wave-packet peak,

$$\left.\frac{\partial\Phi}{\partial\omega}\right|_\theta = 0, \tag{7a}$$

$$\left.\frac{\partial\Phi}{\partial\theta}\right|_\omega = 0. \tag{7b}$$

Substituting Φ into the first equation and using our earlier definition of ϕ_0, we can write an expression for the time $t = \tau_\gamma$ and position $y = \Delta y$ at which the peak first appears in region III,

$$\tau_\gamma = \left.\frac{\partial\phi_0}{\partial\omega}\right|_\theta + \frac{n_1}{c}\Delta y \sin\theta. \tag{8}$$

Equation (8) clearly shows that there are two contributions to the electromagnetic group delay τ_γ. The first term is the frequency derivative analog to (4) and represents the time delay due to the tunneling in the x direction. The second term can be interpreted as the time delay due to the Goos-Hänchen shift Δy, which can be related to the angular dispersion of the transmitted phase $\partial\phi_0/\partial\theta$ by substitution of Φ into (7b),

$$\Delta y = \frac{c}{n_1\omega\cos\theta}\left.\frac{\partial\phi_0}{\partial\theta}\right|_\omega. \tag{9}$$

Steinberg and Chiao then proceed to decompose the partial derivatives $\partial\phi_0/\partial\omega$ and $\partial\phi_0/\partial\theta$ in terms of derivatives in E and V_0 from the 1D problem. By considering cases where $\partial V_0/\partial\omega$ and $\partial E/\partial\omega$ vanish, they find relations between the 2D group delay τ_γ and the 1D group delay τ_g. Rather than reproduce the entire derivation, we simply summarize their results below.

In the low-energy limit $E \ll V_0$, analogous to grazing incidence in the 2D problem, several approximations lead to the relationship

$$\tau_\gamma \approx \frac{n_1^2\hbar\omega}{mc^2}\tau_g.$$

By inspection, if the particle effective mass is chosen such that $mc^2 = n_1^2\hbar\omega$, the delays for the 1D and 2D problems become identical. If instead one considers the "critical" limit ($E \approx V_0$) or the "semiclassical" limit ($E > V_0$) and makes slightly different approximations, the group delay can be expressed as

$$\tau_\gamma \approx \frac{n_2^2\hbar\omega}{mc^2}\tau_g.$$

In [8], the n_2^2 is omitted as that work considers an air gap ($n_2 = 1$). It is also noted that in the semiclassical regime the stationary-phase approximation may break down due to multiple reflections. However, it may remain valid at sufficiently high energies ($E \gg V_0$) or for less abrupt potential differences (e.g., an apodized barrier) where multiple reflections are suppressed or become less important.

III. DWELL TIME AND SELF-INTERFERENCE DELAY

The 1D group delay can also be developed using a variational method outlined in [11,16]. This method makes use of a dwell time τ_d as defined in [5].

$$\tau_d = \frac{\int_0^L |\Psi(x)|^2 dx}{j_{\text{in}}}, \tag{10}$$

where $j_{\text{in}} = \hbar k/m$ is the incident particle flux. While this dwell-time definition is an integral over a stationary state, it can be shown to be equivalent to an integration of the time-dependent wave function $\Psi(x,t)$ (normalized to unity) over the barrier region and over all times [17,18]. Since we have expressions for $\Psi(x)$, C, and D within the barrier, we can evaluate τ_d directly from Eq. (10), yielding

$$\tau_d = \frac{mL}{\hbar k}\frac{\cos^2\phi_0}{2} \times \left[\left(1 + \frac{k^2}{\kappa^2}\right)\frac{\tanh\kappa L}{\kappa L} + \left(1 - \frac{k^2}{\kappa^2}\right)\text{sech}^2\kappa L\right]. \tag{11}$$

It is clear that, while strikingly similar, $\tau_d \neq \tau_g$ derived with the phase-time approach. The relationship between τ_g and τ_d can be calculated from the time-independent Schrödinger equation for $\Psi(x)$ [11], yielding

$$\tau_g - \tau_d = -\hbar\,\text{Im}(R)\frac{\partial}{\partial E}(\ln k) \equiv \tau_i. \tag{12}$$

This extra delay τ_i has been termed the self-interference delay, as it arises out of the overlap of incident and reflected waves in region I. We can put it in a simpler form by evaluating the energy derivative,

$$\tau_i = -\frac{\text{Im}(R)}{k j_{\text{in}}}, \tag{13}$$

which has a form reminiscent of a one-dimensional scattering cross section. Using our calculated expression for R, we can evaluate τ_i explicitly,

$$\tau_i = \frac{mL}{\hbar k}\frac{\cos^2\phi_0}{2}\left(1 + \frac{\kappa^2}{k^2}\right)\frac{\tanh\kappa L}{\kappa L}, \tag{14}$$

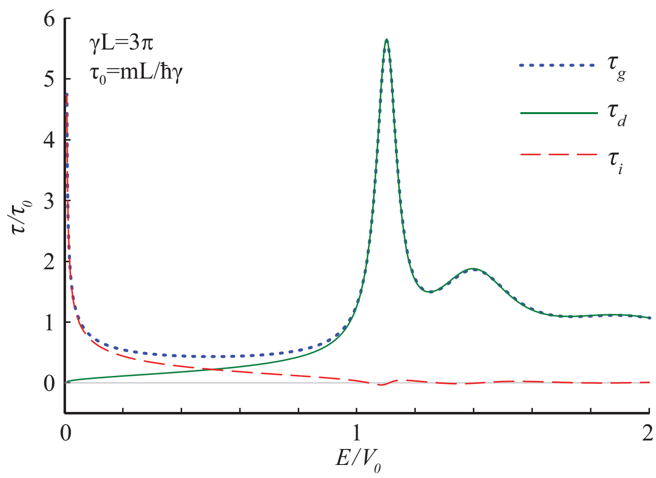

FIG. 3. (Color online) Group delay, dwell time, and self-interference delay as a function of normalized particle energy E/V_0 for a simple one-dimensional barrier. The delays are normalized to the equal time $\tau_0 = L/v_0$, with $v_0 = \hbar\gamma/m$ and $\gamma = \sqrt{2mV_0}/\hbar$. This equal time 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 $\gamma L = 3\pi$.

from which it is clear that the sum of τ_i and τ_d as calculated with the dwell-time definition does indeed equal τ_g calculated with the phase-time method.

Figure 3 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.

In the electromagnetic analog to tunneling, the dwell time is defined by average stored energy and input power rather than number of particles,

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

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, which need not be equal for an arbitrary structure. A similar expression for the self-interference time can be arrived at through the variational theorem,

$$\tau_i = \frac{\langle U_{\text{m}} \rangle - \langle U_{\text{e}} \rangle}{P_{\text{in}}},$$

where U_m and U_e are the magnetic and electric contributions to the time-averaged stored energy. Note that this expression assumes plane-wave propagation in region I; Winful demonstrates this calculation for a one-dimensional waveguide geometry in [6] and finds an additional dispersive factor in the result. In that work, he also demonstrates that these definitions lead to results consistent with the phase-time method. However, to our knowledge, this decomposition into dwell-time and self-interference-delay components has not previously been applied to the 2D tunneling problem.

IV. DWELL TIME DECOMPOSITION OF 2D TUNNELING

The extension into two dimensions introduces an additional complication to the derivation of the tunneling delay. In the FTIR system, the incident light now has two distinct polarization states, 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 authors choose to deal with only the TE case [7,8,10,19], though several address the differences in passing [20] or directly [9]. We have developed a unified derivation that can properly represent either polarization state without difficulty.

To find a solution that is valid for either polarization, one may take the expressions for $\Psi(x,y)$ in (II) and use the TM boundary conditions to solve for C, D, R, and T in terms of the quantities k, κ, n_1, n_2, and L. The necessary modifications can then be identified by comparing these results to those of the TE case. In the following sections, we will summarize this process by comparing the two sets of boundary conditions and presenting a modified notation that successfully satisfies either set of boundary conditions.

A. TE boundary conditions

For an incident TE wave, the electric fields involved are

$$\mathbf{E}_{\text{I}} = E_0(e^{ik_x x} + Re^{-ik_x x})e^{ik_y y - i\omega t}\hat{\mathbf{z}}, \qquad (15a)$$

$$\mathbf{E}_{\text{II}} = E_0(Ce^{-\kappa x} + De^{\kappa x})e^{ik_y y - i\omega t}\hat{\mathbf{z}}, \qquad (15b)$$

$$\mathbf{E}_{\text{III}} = E_0(Te^{ik_x x})e^{ik_y y - i\omega t}\hat{\mathbf{z}}, \qquad (15c)$$

where again we will refer to the components in parentheses as ψ. The boundary conditions are obtained by enforcing continuity of the transverse components of $\mathbf{E}$ and $\mathbf{H}$, which for the TE case are E_z and H_y, at $x = 0$ and $x = L$. Continuity of E_z dictates that ψ is continuous, while continuity of H_y requires continuity of $\partial\psi/\partial x$. These four conditions are identical to the boundary conditions on Ψ for the one-dimensional problem and successfully demonstrate a one-to-one mapping of the two-dimensional problem to the one-dimensional case. Thus, the solutions presented earlier in the paper for the one-dimensional problem also successfully address the 2D TE problem.

B. 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 a similar fashion to the TE case,

$$\mathbf{H}_{\text{I}} = H_0(e^{ik_x x} + R_{\text{m}}e^{-ik_x x})e^{ik_y y - i\omega t}\hat{\mathbf{z}}, \qquad (16a)$$

$$\mathbf{H}_{\text{II}} = H_0(C_{\text{m}}e^{-\kappa x} + D_{\text{m}}e^{\kappa x})e^{ik_y y - i\omega t}\hat{\mathbf{z}}, \qquad (16b)$$

$$\mathbf{H}_{\text{III}} = H_0(T_{\text{m}}e^{ik_x x})e^{ik_y y - i\omega t}\hat{\mathbf{z}}. \qquad (16c)$$

The boundary conditions are still derived from the continuity of all transverse components of **E** and **H**, but in the TM case these components are H_z and E_y. Continuity of H_z again enforces continuity of ψ. However, in the TM case the continuity of E_y enforces the continuity of $(1/\epsilon)(\partial\psi/\partial x)$ at each interface, or equivalently $(1/n^2)(\partial\psi/\partial x)$ for a lossless material. While the conversion from ϵ to n^2 implicitly assumes that the materials in all regions are lossless, the results should retain validity for very weakly lossy materials.

The difference in boundary conditions prevents us from using the results presented earlier to properly characterize the TM case. However, we will now show that with a slight modification to Eqs. (2) and (3), we can produce a solution that satisfies both TM and TE boundary conditions.

C. Generalized solution

This additional complication in boundary conditions can be easily accommodated with the introduction of a factor M_{ij},

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

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 $M_{12} = 1/M_{21}$. This definition of M_{ij} is equivalent to the factor m found in [20,21] or the encapsulation of n_1 and n_2 into the particle effective mass in [9].

If we evaluate the boundary condition expressions for the TM case, we find the following expressions for Δ_m, Δ'_m, g_m, R_m, T_m, C_m, and D_m, where the subscript m is used to clearly differentiate these quantities from their one-dimensional counterparts:

$$\Delta_\mathrm{m} = \frac{1}{2}\left(\frac{M_{12}\kappa}{k_x} - \frac{k_x}{M_{12}\kappa}\right), \tag{17a}$$

$$\Delta'_\mathrm{m} = \frac{1}{2}\left(\frac{M_{12}\kappa}{k_x} + \frac{k_x}{M_{12}\kappa}\right), \tag{17b}$$

$$g_\mathrm{m} = \cosh\kappa L + i\Delta_\mathrm{m}\sinh\kappa L, \tag{17c}$$

$$R_\mathrm{m} = -\frac{i\Delta'_\mathrm{m}}{g_\mathrm{m}}\sinh\kappa L, \tag{17d}$$

$$T_\mathrm{m} = \frac{e^{-ik_xL}}{g_\mathrm{m}}, \tag{17e}$$

$$C_\mathrm{m} = \left(1 - i\frac{k_x}{M_{12}\kappa}\right)e^{\kappa L}/2g_\mathrm{m}, \tag{17f}$$

$$D_\mathrm{m} = \left(1 + i\frac{k_x}{M_{12}\kappa}\right)e^{-\kappa L}/2g_\mathrm{m}. \tag{17g}$$

The expressions for R_m, T_m, and the factor g_m are unchanged from the one-dimensional case apart from using the updated definitions of Δ_m and Δ'_m. The expressions for C_m, D_m, Δ_m, and Δ'_m now include factors of M_{12}, though they simplify to the old expressions when $M_{12} \to 1$ (as for the TE case).

With these redefinitions, we now have a general set of expressions for R_m, C_m, D_m, and T_m based on the boundary conditions of our two-dimensional tunneling problem that is valid for both TM and TE situations. 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. We can now proceed with explicit calculation of the dwell time, the self-interference delay, and the group delay for 2D tunneling. 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.

D. Dwell time

The general definition of dwell time from Eq. (10) is, in the notation of our two-dimensional problem,

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

where $j_\mathrm{in} = c/n_1$ remains the incident particle flux and $\psi(x)$ now replaces the one-dimensional wave function $\Psi(x)$. Because of the translational symmetry 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 do k_x and κ.

If we evaluate this integral using the expression for ψ in region II and substitute our expressions for C_m and D_m into the result, we can put the expression for the dwell time in a more recognizable form.

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

Note that this is the same expression we have for τ_d in Eq. (11), but with additional factors of M_{12} to support two-dimensional TM boundary conditions. In fact, we can write both τ_d and τ_δ in a slightly simpler form by using our definitions of Δ, Δ', Δ_m, and Δ'_m and the 1D particle flux $j_\mathrm{in}^{(1\mathrm{D})} = \hbar k/m$:

$$\tau_d = \frac{L\cos^2\phi_0}{j_\mathrm{in}^{(1\mathrm{D})}}\frac{k}{\kappa}\left[\Delta'\frac{\tanh\kappa L}{\kappa L} + \Delta\mathrm{sech}^2\kappa L\right],$$

$$\tau_\delta = \frac{L\cos^2\phi_0}{j_\mathrm{in}M_{12}}\frac{k_x}{\kappa}\left[\Delta'_\mathrm{m}\frac{\tanh\kappa L}{\kappa L} + \Delta_\mathrm{m}\mathrm{sech}^2\kappa L\right].$$

So the definition of dwell time carries over very well to the two-dimensional electromagnetic case. The only differences are an additional factor of $1/M_{12}$, the replacement of Δ and Δ' with Δ_m and Δ'_m, respectively, and of course a different definition of j_in than in the 1D particle case. If we approximate $\Delta_\mathrm{m} \approx \Delta$ and $\Delta'_\mathrm{m} \approx \Delta'$, which is valid when M_{12} is close to 1, we can relate τ_δ to τ_d,

$$\tau_\delta = \frac{j_\mathrm{in}^{(1\mathrm{D})}}{M_{12}j_\mathrm{in}}\tau_d. \tag{18}$$

Note that while this equation requires the mentioned approximation for the TM case, it is identically true for the TE case since $M_{12} = 1$.

E. Self-interference delay

To express τ_ι, we start with the definition in Eq. (13) with k_x substituted for k,

$$\tau_\iota = -\frac{\mathrm{Im}(R_\mathrm{m})}{k_x j_\mathrm{in}}.$$

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 energy derivative in (12) could be represented as a frequency derivative in the electromagnetic case and leads to a factor of $1/v_\mathrm{group}$. Since our materials are only weakly dispersive for the photon bandwidths we are interested in, we can approximate this as n/c, or $1/v_\mathrm{phase}$. Substitution of $\mathrm{Im}(R_\mathrm{m}) = -\Delta'_\mathrm{m}\cos^2\phi_0 \tanh\kappa L$ yields

$$\tau_\iota = \frac{L}{j_\mathrm{in}}\frac{\cos^2\phi_0}{2}\left(1+\frac{\kappa^2}{k_x^2}\right)\frac{\tanh\kappa L}{\kappa L},$$

which is identical to Eq. (14). In this case, the only difference between τ_i and τ_ι is the choice of j_in such that

$$\tau_\iota = \frac{j_\mathrm{in}^{(1\mathrm{D})}}{j_\mathrm{in}}\tau_i, \tag{19}$$

using the same definitions of j_in and $j_\mathrm{in}^{(1\mathrm{D})}$ as in (18).

F. Group delay

The phase of the transmitted field at $x = L$ is still $\Phi = \phi_t + k_x L + k_y y$ as before. Any variation due to the polarization state is encapsulated in ϕ_t. If we evaluate Eqs. (7) and eliminate y, we are left with

$$\tau_\gamma = \left(\frac{\partial\phi_0}{\partial\omega}\right)_\theta - \frac{\tan\theta}{\omega}\left(\frac{\partial\phi_0}{\partial\theta}\right)_\omega. \tag{20}$$

We desire to evaluate these partial derivatives of ϕ_0 in order 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 in terms of k_x and κ derivatives. This process is algebraically tedious but straightforward; as such we will only summarize the results.

$$\begin{aligned}\left(\frac{\partial\phi_0}{\partial\theta}\right)_\omega &= \left(\frac{\partial\phi_0}{\partial\kappa}\right)_{k_x}\left(\frac{\partial\kappa}{\partial\theta}\right)_\omega + \left(\frac{\partial\phi_0}{\partial k_x}\right)_\kappa\left(\frac{\partial k_x}{\partial\theta}\right)_\omega\\ &= -k_x j_\mathrm{in}\tan\theta(M_{12}\tau_\delta + \tau_\iota).\end{aligned}$$

$$\begin{aligned}\left(\frac{\partial\phi_0}{\partial\omega}\right)_\theta &= \left(\frac{\partial\phi_0}{\partial\kappa}\right)_{k_x}\left(\frac{\partial\kappa}{\partial\omega}\right)_\theta + \left(\frac{\partial\phi_0}{\partial k_x}\right)_\kappa\left(\frac{\partial k_x}{\partial\omega}\right)_\theta\\ &= -\frac{\kappa^2 j_\mathrm{in}}{k_x\omega}M_{12}\tau_\delta + \frac{k_x j_\mathrm{in}}{\omega}\tau_\iota.\end{aligned}$$

Substituting these into (20) and simplifying, we get our expression for τ_γ,

$$\tau_\gamma = \frac{j_\mathrm{in}\omega}{k_x c^2}\left[M_{12}n_2^2\tau_\delta + n_1^2\tau_\iota\right]. \tag{21}$$

This result demonstrates the decomposition of the group delay τ_γ into two components, one proportional to the dwell time τ_δ and one proportional to the self-interference delay τ_ι. We can relate τ_γ directly to τ_d and τ_i through Eqs. (18) and (19). If we make those substitutions along with the particle flux $j_\mathrm{in}^{(1\mathrm{D})} = \hbar k_x/m$, we find

$$\tau_\gamma = \frac{\hbar\omega}{mc^2}\left[n_2^2\tau_d + n_1^2\tau_i\right]. \tag{22}$$

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, k, and θ in Table I,

$$V_0 = \frac{\hbar^2\omega^2}{2mc^2}\left(n_1^2 - n_2^2\right), \tag{23a}$$

$$E = V_0\frac{n_1^2\cos\theta}{n_1^2 - n_2^2}. \tag{23b}$$

Note that due to the approximation made in (18), Eq. (22) is only strictly valid for the TE geometry. In the TM geometry, it provides accurate results as long as $\tau_d \leqslant \tau_i$ and exhibits the same qualitative behavior as the TE case for any choice of input parameters but begins to become numerically inaccurate when τ_d is the dominant delay contribution.

As a final confirmation, let us compare our results to those found in Steinberg and Chiao's analysis of this problem in [8]. In Eq. (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 (23a) and (23b) 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 (22) simplifies to

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

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 in the Wentzel-Kramers-Brillouin (WKB) or semiclassical limit, where $E > V_0$ and κ becomes imaginary. Under either of those conditions, (22) simplifies to

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

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 semiclassical limits that effective mass is reduced to $mc^2 = \hbar\omega$. Our expression clarifies this difference 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 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 semiclassical 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, Eq. (22) could be interpreted as an energy-weighted average delay, in which each delay term is weighted by the permittivity of that region to properly account for the difference in energy density between a massive particle and electromagnetic waves.

It is interesting that both the TE and TM cases map to the one-dimensional problem according to (22), 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 $j_{\rm in} = c/n_1$ into Eq. (21) to find

$$\tau_\gamma = \left[M_{12}\frac{n_2^2}{n_1^2}\tau_\delta + \tau_\iota\right].$$

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 as long as τ_δ is significant compared to τ_ι. This results appears consistent with the delays measured in [22].

V. FTIR IN A DOUBLE-PRISM GEOMETRY

The theoretical treatment presented in the previous section addressed the general case of tunneling delay in a glass-dielectric-glass geometry and is consistent with the usual treatment of the topic [7–9]. However, this treatment assumes the glass regions are infinite in transverse extent, which is not representative of experimental reality. The usual experimental implementation mimics that proposed in [8,10], which is a system of two prisms separated by an air gap. The theory presented in the literature thus far does not accurately reflect the experimental measurements one can obtain in this system because the practical issue of coupling into and out of the double-prism system is generally overlooked or ignored.

To illustrate this discrepancy, we present Fig. 4, which shows a simple diagram of total internal reflection (TIR) at

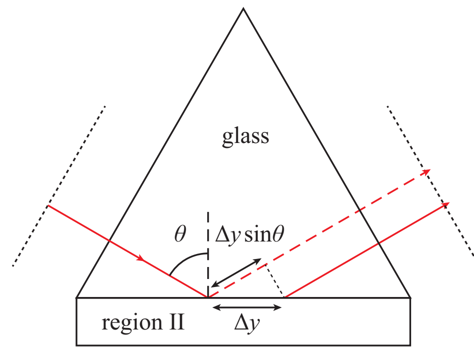

FIG. 4. (Color online) TIR at the interface between an equilateral prism and a dielectric slab. The Goos-Hänchen shift Δy causes a change in the glass propagation length, which has ramifications for an experimental measurement of tunneling delay in the double-prism geometry.

the interface between an equilateral prism and a slab of an arbitrary dielectric or conducting material. If the second region were filled with a perfectly conducting material, the light would reflect along the dashed red line, with no Goos-Hänchen shift Δy. However, if region II is air, we observe TIR with a nonzero Goos-Hänchen shift, and the reflected beam propagates through a smaller length of glass. The difference in propagation distance is simply $\Delta y \sin\theta$. Thus, the delay we measure experimentally in an FTIR configuration is not the τ_γ derived in the previous section and in the bulk of the literature, but

$$\tau_{\gamma,\rm meas} = \tau_\gamma - n\Delta y \sin\theta/c. \tag{24}$$

By inspection of Eq. (8), we see that

$$\tau_{\gamma,\rm meas} = \left(\frac{\partial\phi_0}{\partial\omega}\right)_\theta. \tag{25}$$

As discussed in Sec. II, the term $\partial\phi_0/\partial\omega$ describes the delay contribution from the propagation in the x direction, while the suppressed term $\tau_{\gamma,\rm unmeas} = n\Delta y \sin\theta/c$ describes the contribution from the y direction, or Goos-Hänchen shift. This is perhaps more clearly seen by considering the explicit form of this partial derivative, as derived earlier:

$$\tau_{\gamma,\rm meas} = \frac{j_{\rm in}}{k_x\omega}\left[k_x^2\tau_\iota - \kappa^2 M_{12}\tau_\delta\right]. \tag{26}$$

Note that below the critical angle $\kappa = ik_{2x}$ and $-\kappa^2 = k_{2x}^2$, where we have 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,

$$\tau_{\gamma,\rm unmeas} = \frac{j_{\rm in}}{k_x\omega}\left[k_y^2(\tau_\iota + M_{12}\tau_\delta)\right]. \tag{27}$$

We can also express the measurable and unmeasurable portions in terms of the one-dimensional delays τ_d and τ_i using Eqs. (18) and (19).

$$\tau_{\gamma,\rm meas} = \frac{\hbar}{m\omega}\left[k_x^2\tau_i - \kappa^2\tau_d\right], \tag{28a}$$

$$\tau_{\gamma,\rm unmeas} = \frac{\hbar}{m\omega}\left[k_y^2(\tau_i + \tau_d)\right]. \tag{28b}$$

In other words, apart from a constant factor, the "measurable" portion of the delay is a weighted average of the dwell and self-interference delays, with each component weighted by the x component of the wave vector in the appropriate region. The unmeasurable portion is the complementary expression, the average as weighted by the y component of the wave vector, which is the same in each region. If we assume normal incidence (Fabry-Pérot operation) and substitute $k_x = n_1\omega/c$ and $\kappa = in_2\omega/c$ into Eq. (28), $\tau_{\gamma,\mathrm{meas}}$ bears a striking resemblance to Eq. (22). 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-Pérot regime.

Equations (25)–(27) 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 compensated by a reduction in glass propagation length. 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 polygon. 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 wave vectors.

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 a realistic device, τ_γ may be on the order of picoseconds at a Fabry-Pérot resonance, but $\tau_{\gamma,\mathrm{meas}}$ is smaller by a factor of 100 [23]. Similar discrepancies can be observed in the barrier region.

This practical issue has been overlooked in the majority of the literature [7,9,10,24]. 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" [25]. 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.

Rather than fault this result as a failing of the double-prism system, we may 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 evanescent (i.e., "tunneling") aspects of FTIR and thus more intimately observe its relationship to the one-dimensional problem.

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 [22,26].

VI. CONCLUSIONS

In this paper, we have recast the traditional two-dimensional FTIR problem in terms of a cavity model, with the total group delay divided into self-interference and dwell-time terms. Once expressed in these terms, the 2D 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 formal similarity between the time-independent Schrödinger equation and the Helmholtz equation. Furthermore, we have shown that our more general version simplifies to match previous predictions in the special cases considered by Steinberg and Chiao in [8].

In addition, this interpretation of the tunneling delay phenomenon gives us physical insight into the process. Conceptually, it breaks the process down into delays due to cavity-like effects in the tunneling region and 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.

Finally, we have identified a peculiarity of the double-prism FTIR system that has not yet 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 $\tau_{\gamma,\mathrm{meas}}$ and τ_γ that 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.

ACKNOWLEDGMENTS

We acknowledge many useful discussions with Herbert Winful. This work was supported by the US Defense Threat Reduction Agency–Joint Science and Technology Office for Chemical and Biological Defense (Grant No. HDTRA1-10-1-0025) and by the Canada Excellence Research Chairs Program.

[1] L. MacColl, Phys. Rev. **40**, 621 (1932).
[2] E. Wigner, Phys. Rev. **98**, 145 (1955).
[3] T. E. Hartman, J. Appl. Phys. **33**, 3427 (1962).
[4] A. M. Steinberg, P. G. Kwiat, and R. Y. Chiao, Phys. Rev. Lett. **71**, 708 (1993).
[5] M. Büttiker, Phys. Rev. B **27**, 6178 (1983).
[6] H. G. Winful, Phys. Rev. E **68**, 016615 (2003).
[7] A. Ghatak and S. Banerjee, Appl. Opt. **28**, 1960 (1989).
[8] A. M. Steinberg and R. Y. Chiao, Phys. Rev. A **49**, 3283 (1994).
[9] B. Lee and W. Lee, J. Opt. Soc. Am. B **14**, 777 (1997).

[10] R. Chiao, P. Kwiat, and A. M. Steinberg, Phys. B **175**, 257 (1991).
[11] H. G. Winful, Phys. Rev. Lett. **91**, 260401 (2003).
[12] H. Winful, Phys. Rep. **436**, 1 (2006).
[13] M. Y. Azbel, Phys. Rev. B **28**, 4106 (1983).
[14] J. P. Falck and E. H. Hauge, Phys. Rev. B **38**, 3287 (1988).
[15] A. M. Steinberg, Phys. Rev. Lett. **74**, 2405 (1995).
[16] F. Smith, Phys. Rev. **118**, 349 (1960).
[17] C. R. Leavens and G. C. Aers, Phys. Rev. B **39**, 1202 (1989).
[18] E. H. Hauge, J. P. Falck, and T. A. Fjeldly, Phys. Rev. B **36**, 4203 (1987).
[19] A. B. Shvartsburg, Phys. Usp. **50**, 37 (2007).
[20] J. C. Martinez and E. Polatdemir, Phys. Lett. A **351**, 31 (2006).
[21] J. C. Martinez, E. Polatdemir, and M. Taufik, Eur. Phys. J. D **41**, 49 (2006).
[22] P. Balcou and L. Dutriaux, Phys. Rev. Lett. **78**, 851 (1997).
[23] G. M. Gehring, Ph.D. thesis, University of Rochester, 2011.
[24] D. J. Papoular, P. Clade, S. V. Polyakov, C. F. McCormick, A. L. Migdall, and P. D. Lett, Opt. Express **16**, 16005 (2008).
[25] A. Haibel and G. Nimtz, Ann. Phys. (NY) **10**, 707 (2001).
[26] M. Deutsch and J. E. Golub, Phys. Rev. A **53**, 434 (1996).