Tunneling delay times in one and two dimensions

Aephraim M. Steinberg and Raymond Y. Chiao
Department of Physics, University of California at Berkeley, Berkeley, California 94720
(Received 18 May 1993)

We demonstrate that although the well-known analogy between the time-independent solutions for two-dimensional tunneling (e.g., frustrated total internal reflection) and tunneling through a one-dimensional potential barrier cannot, in general, be extended to the time domain, there are certain limits in which the delay times for the two problems obey a simple relationship. In particular, when an effective mass is chosen such that $mc^2=\hbar\omega$, the "classical" traversal times for allowed transmission become identical for a photon of energy $\hbar\omega$ traversing an air gap between regions of index n and for a particle of mass m traversing the analogous square barrier of height V_0 in one dimension. The quantum-mechanical group delays are also identical, given this effective mass, both for $E\approx V_0$ ($\theta\approx\theta_c$) and for $E\gg V_0$ ($\theta\ll\theta_c$). (For a smoothly varying potential or index of refraction, the agreement persists for all values of E where the WKB approximation applies.) The same relation serves to equate the quantum-mechanical "dwell" times for any values of E and V_0. On the other hand, in the "deep tunneling" limit, $E\ll V_0$ ($\theta\approx\pi/2$), one must choose $mc^2=n^2\hbar\omega$ in order to make the group delays equal for the two problems. These equivalences simplify certain calculations, and the two-dimensional analogy may also be useful for geometrically visualizing the tunneling process and the anomalously small group delays known to occur in the opaque limit. We also demonstrate that the equality of the group delays for transmission and reflection for lossless barriers follows from a simple intuitive argument based on time-reversal invariance, and discuss the extension of the result to the case of lossy barriers.

PACS number(s): 03.65.Bz, 42.25.Bs, 73.40.Gk

I. INTRODUCTION

Over the past few years, renewed attention has been devoted to the long-standing controversy over the duration of the tunneling process [1–9]. Most of the theoretical work has centered on electron tunneling in one dimension, although some recent papers [10–14] have found it advantageous to focus on electromagnetic instances of tunneling. Both one- and two-dimensional tunneling are important in solid-state physics as well as in electromagnetism, and analogies between the different processes can be fruitful [15]. Such analogies are well known for the time-independent problems, but are nontrivial for the time-dependent case. Here we discuss the extent to which the dynamics of one- and two-dimensional tunneling (as well as allowed transmission) can be considered analogous, and present some results about the relationship of the reflection and transmission times for tunneling (regardless of dimensionality).

A. Group-delay times

While there have been several definitions offered for tunneling times, we are going to discuss primarily the group-delay prediction, also known as the "phase time." This quantity is based on the stationary-phase approximation, and is intended to give the time at which the peak of a tunneling wave packet will appear at the far side of a barrier, relative to the time at which the extrapolated peak of the incident packet would reach the entrance face of the barrier [16]. (To avoid confusion with the phase velocity, we use the terminology "group-delay time," or simply "delay time"; when necessary, we make a distinction between "transmission delay times" and "reflection delay times," and for brevity in those cases we have on occasion dropped the word "delay." This should not be mistaken for an introduction of an alternate, undefined quantity.) In some limits, this time may be less than the thickness of the barrier divided by the vacuum velocity of light c, which has led researchers to question this approach, both mathematically and interpretationally [2,5,6]. Indeed, it has been stated that when the group velocity exceeds c, it "is just not a useful concept" [17] and that "physics has no law about a peak turning into a peak" [18]. (For example, the transmitted pulse might be so distorted that the method of stationary phase has no physical content; this can be true, for instance, in the classical limit of allowed transmission, where multiple reflections are displaced relative to one another rather than interfering; see discussion in Sec. III.) We proposed an optical experiment to test this prediction at the single-particle level. We showed that in our experiment, the group delay was indeed a meaningful quantity, in that the distortion of the wave packet was sufficiently slight as to justify the identification of the peak as a robust characteristic. More importantly, it correctly described the tunneling delay time at the single-photon level [14], even

1050-2947/94/49(5)/3283(13)/$06.00

though the apparent tunneling velocity exceeded the vacuum speed of light. By using pairs of "conjugate" particles emitted essentially simultaneously, allowing only one to tunnel, and then comparing their arrival times, we provided an unambiguous operational definition of the tunneling delay. We compared the particles' arrival times before and after inserting a barrier in the path of one of them; thus we measured the difference between the tunneling delay time as defined above and the delay time for traversing free space, simply d/c for a photon. The other proposed definitions for tunneling times may well be important in different types of experiments; in this paper, however, we concern ourselves only with the delay.

Our original proposal [12] relied on the process of frustrated total internal reflection (FTIR), in which a photon incident on a planar glass-air interface at an angle greater than the critical angle for total reflection may "tunnel" through an air gap of width d to a second glass region, parallel to the first (see Fig. 1). We showed that the solutions to Maxwell's equations for this problem could have exactly the same form as the solutions to Schrödinger's equation for an electron incident upon a one-dimensional square barrier of the same width d. In particular, for s-polarized light, the matching conditions for the electric field $\mathcal{E}$ are the same as those for the wave function Ψ (viz., both the field and its derivative are continuous). Due to translation symmetry, the y dependence in the two-dimensional problem must be the same in all three regions, and can thus be dropped. Setting both the x component k_x of the incident wave vector and the evanescent decay constant (in the barrier region) κ to be the same in the one-dimensional Schrödinger and two-dimensional Maxwell problems leads to identical transmission amplitudes for the two problems. This is accomplished when the equivalences

$$\frac{2mE}{\hbar^2} \Longleftrightarrow \frac{\omega^2}{c^2} n^2 \cos^2\theta \, , \qquad \frac{2m(V_0 - E)}{\hbar^2} \Longleftrightarrow \frac{\omega^2}{c^2}(n^2 \sin^2\theta - 1) \tag{1}$$

are drawn. (Here and throughout, m is the mass of the electron, E is its energy, and V_0 is the height of the square barrier; ω is the frequency of the photon, n is the index of refraction in the two glass regions, and θ is the angle of incidence of the photon.) To this kinematic identification, we were not immediately able to add a dynamical identification. That is, while the transmission probabilities have the same form for the time-independent problems, the extension to the time-dependent case is nontrivial. There are two reasons for this. First of all, our proposed experiment involved photons, whose dispersion relation is different from that of the massive particles implicit in the Schrödinger equation. (Note that in the above equations, the frequency of the Schrödinger stationary states, $E/\hbar$, is not the same as the frequency ω of the corresponding optical stationary states, although the wave vector k_x is the same in both problems.) The second difficulty arises because in a two-dimensional experiment, there may be a transverse (parallel to the air-glass interfaces) shift Δy for the transmitted beam. This would occasion an additional phase shift of $e^{ik_y \Delta y}$, apart from the phase shift of the transmission amplitude itself. For these two reasons, we originally attempted no mathematical identification of the traversal times for the two problems, but planned simply to test the group delay in the optical (two-dimensional) problem, allowing arguments to be made by qualitative analogy for the one-dimensional case.

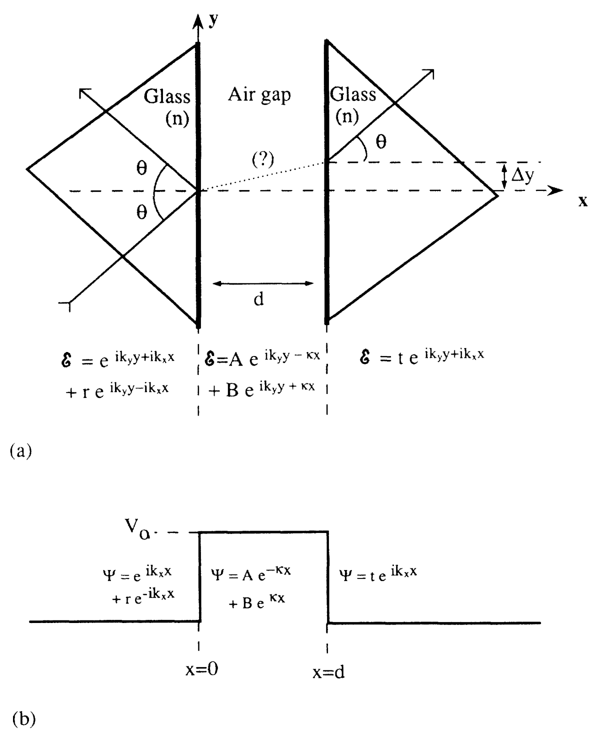

FIG. 1. (a) Two-dimensional tunneling (FTIR) through an air gap when $\theta > \theta_c = \sin^{-1}(1/n)$. (b) One-dimensional tunneling for $E < V_0$. If k_x and κ are arranged to be the same as in (a), then the transmission and reflection coefficients t and r will be the same in the two problems.

B. One- and two-dimensional experiments

The difference between the one- and two-dimensional problems is particularly apparent near criticality, that is, for $E \approx V_0$ or $\theta \approx \theta_c \equiv \sin^{-1}(1/n)$. In the one-dimensional problem, the physicality of the group delay may be questioned in this regime for the following reason. In order to perform an accurate time measurement, one must construct a finite wave packet consisting of many frequency components. As $E \rightarrow V_0$, however, these components will have very different tunneling characteristics, and the high-energy tails will even extend into the region of allowed transmission. The resulting distortion of the pulse may make the group delay meaningless. In the two-dimensional case, by contrast, we see from (1) that the role of the critical parameter is played not by ω but rather by θ; that is, an arbitrarily short pulse could be incident on the barrier, with every component just "below the barrier." As we pointed out in [12], this should allow a regime of tunneling times to be studied in FTIR which would be experimentally inaccessible (and perhaps theoretically questionable) in the analogous one-dimensional case.

The above is a slight oversimplification. In fact, for θ to be perfectly well defined implies an unbounded plane wave. For a plane-wave pulse incident on a barrier at an angle, the group delay is a quantity of dubious significance, regardless of how short the pulse; the infinite transverse extent of the pulse means that even if it reaches the origin at $t=0$, it has been interacting with the barrier since $t=-\infty$ (far down along the y axis). Thus a two-dimensionally bounded wave packet must instead by considered, implying some spread in θ, analogous to the spread in E discussed earlier. At an experimental level, this is still preferable to the one-dimensional case. While in the latter, the smaller the spread in E the worse the time resolution, in the two-dimensional case the time resolution is determined principally by ω, not by θ; while some spread in θ is unavoidable, it need not be so large as to smear out the tunneling characteristics. (A 2-mm-wide pulse, for example, can be constructed with a duration up to about 1000 times smaller than its transverse width.)

Despite the advantages of FTIR, the one-dimensional problem is of more current interest, and is somewhat less complicated at a technical level. For these reasons, before embarking on the proposed experiment, we performed a simpler one [14], in which we confirmed that a single-photon wave packet tunneling through a one-dimensional barrier appeared essentially undistorted at the far side of the barrier, at a time described by the group delay, even though this implied an effective mean velocity greater than c. In our experiment, the barrier was a multi- (11-) layer dielectric mirror, which can be seen as an optical realization of a Kronig-Penney band-gap material [19]. The width of the barrier could not be adjusted. Because there are some questions about the identification of such a medium with a tunnel barrier, and since one of the most striking predictions is that the group delay becomes independent of the barrier thickness for opaque barriers (i.e., in the limit of low transmission), we plan to perform the FTIR experiment as well. The size of the FTIR air gap is adjustable, and the wave in this region is purely evanescent.

Evanescent phenomena in two dimensions are most familiar in optics, where they are of great importance and have been extensively studied [20]. Evanescent fields have been quantized [21], the strange behavior of the phase shifts in FTIR has been observed [22], and theoretical papers have discussed the resulting anomalously short group delays [23,24]. However, we wish to stress that two-dimensional tunneling is not exclusively an optical problem, and that in principle, heterostructures could be constructed which would allow the process of frustrated total internal reflection to occur for electrons. In particular, at a linear or planar interface where the potential energy of an electron changes, it is "refracted" as by an effective index of refraction n_{eff} proportional to the square root of its kinetic energy [15,25]. (Effective mass changes may also lead to refractive effects.) For incidence angles greater than the critical value of $\sin^{-1}(1/n_{\mathrm{eff}})$, the electron will undergo total internal reflection. If the region of low effective index is a thin one, and is followed by a second region of high effective index, then some transmission is possible regardless of the angle of incidence. This problem is exactly analogous to the optical case of FTIR.

Conversely, one-dimensional tunneling exists in electromagnetism as well. Microwaves beyond the cutoff frequency of a constricted waveguide may tunnel in one dimension, in a process mathematically equivalent to the tunneling of an electron across a square potential barrier [10,11]. In this paper, however, we will discuss FTIR purely in terms of photons, and one-dimensional tunneling purely in terms of electrons. We do this on the one hand for concreteness and to avoid confusion between the various parameters involved in the two problems, and on the other hand because FTIR is most frequently discussed as an optics phenomenon, while one-dimensional tunneling is usually discussed in the context of the Schrödinger equation.

In addition to simplifying some calculations, the fact that analogies can be drawn between the dynamics of one- and two-dimensional tunneling processes may be useful in attempts to understand the qualitative features of these processes. For example, the "classical" particle traversal time in one dimension diverges as $E \to V_0$ from above, since the velocity in the barrier region tends toward zero. In two dimensions, the same divergence arises not because of any change in the photon velocity, but because its propagation direction becomes more and more parallel to the interfaces. Of course, the group delay remains finite in both cases; while this is clearly connected to the uncertainty principle, it is difficult to visualize in the one-dimensional problem. In two dimensions, by contrast, this effect manifests itself through diffraction. It is not possible for a transversely bounded beam to propagate indefinitely parallel to the interface without eventually diffracting, and thus at least a part of the beam reaches the exit face in some finite time. It is even tempting to argue that the bulk of this transmitted portion originates closer to the right-hand side of the incident beam (see Fig. 2), thus offering a partial explanation of the anomalously small peak delay. The peak of the transmitted wave packet seems to originate earlier in time than the incident peak [26,27], thus leading to a superluminal delay *without* implying that any individual portion of the wave ever traveled at a superluminal speed.

C. Outline

In analyzing our proposed two-dimensional, optical experiment, we have now discovered that in addition to the qualitative analogy, there is in fact a suggestive quantitative relationship between the dynamics of this process and those of one-dimensional tunneling of a massive particle. When the mass of the particle tunneling in one dimension and the energy of the (massless) particle tunneling in two dimensions satisfy the relationship $mc^2=\hbar\omega$, the "dwell" times are the same for the two problems, and the group-delay times are essentially identical both near criticality and for the WKB limit (slowly varying poten-

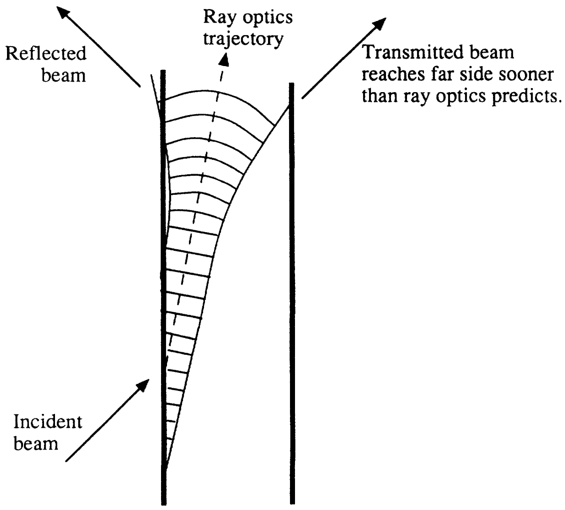

FIG. 2. A simplified picture of the wave nature of transmission through an air gap. When light is incident just below the critical angle, the ray inside the barrier region is nearly parallel to the interfaces; *at* the critical angle, the ray optics traversal time thus diverges. The wave-mechanical group delay, on the other hand, remains finite: the transversely bounded beam undergoes diffraction, and begins to couple out of the barrier sooner than predicted by ray optics. Depending on the various length scales (i.e., how much diffraction occurs), the transmitted portion may preferentially consist of wave vectors closer to the normal or of those parts of the beam which originate closer to the right-hand side and thus reached the barrier earliest. The former case simply corresponds to the preferential transmission of those components which were incident furthest from the critical angle, while the latter may offer some insight into anomalously small group delays. When the angle of incidence exceeds the critical angle, *each* incident k vector corresponds to an evanescent wave in the barrier region, whose Fourier decomposition includes many different angles, with a k-space distribution of width κ.

tial, or high energy). This relationship fails in the tunneling regime, but for "deep tunneling," the low-energy limit, the two group delays become identical if the parameters satisfy $mc^2=n^2\hbar\omega$ instead. Section II will demonstrate these mathematical correspondences. Sections III and IV will treat the connection between group delays and transverse shifts for transmission and for reflection, and briefly discuss the implications for understanding certain features of both. A summary is given in Sec. V, followed by an Appendix which explicitly verifies the equivalence of the group delays in the "opaque" limit.

II. THE CORRESPONDENCE

It is instructive to examine the correspondences of (1) in the quasiclassical or WKB limit. For the electron, this corresponds to a classical particle model, and for the photon, to ray optics (short-wavelength limit). In the former case, when the energy E of the incident electron exceeds the height V_0 of the barrier, the velocity in the barrier is simply $\sqrt{2(E-V_0)/m}$, and thus the traversal time

$$(\tau_e)_{\rm WKB}=d/\sqrt{2(E-V_0)/m}\ . \tag{2}$$

When we apply Eq. (1), we find that this is equal to

$$(\tau_e)_{\rm WKB}=d\Big/\left[\frac{\hbar^2\omega^2}{m^2c^2}(1-n^2\sin^2\theta)\right]^{1/2}=\frac{mc^2}{\hbar\omega}\frac{d/c}{\cos\theta'}\ , \tag{3}$$

where we have defined $\sin\theta'=n\sin\theta$ according to Snell's law. Remarkably, this proportionality is exactly what one expects for the time delay $(\tau_\gamma)_{\rm WKB}$ for a beam of light which is refracted and travels through the air gap at an angle θ' to the normal with a velocity c (see Fig. 3); if we set

$$mc^2=\hbar\omega\ , \tag{4}$$

the two times are identical.

To recap, then, given the mass m of a particle and the height V_0 of a barrier, we use (1) and (4) to uniquely determine the analogous parameters for the optical problem, that is, the frequency ω and the index n. Each incident energy E corresponds to a particular incident angle θ [also from (1)], in such a way that not only the transmission probability but also the delay time is the same for the two problems, at least in the classical limit. This seeming coincidence led us to ask whether the behavior of the two traversal times remains the same when wave-mechanical effects are taken into account.

A. Connection with the "dwell time"

Before calculating the quantum-mechanical group delay, it may be interesting to note that (4) does serve to make the "dwell time" equal for the two problems, in all limits. The dwell times is an interaction time defined in terms of the steady-state problem and without regard for the particular scattering channel (reflection or transmission) taken by a given particle. Specifically, it can be expressed as the ratio of the probability (energy) density within the barrier to the incident probability (energy) flux,

$$\tau_{d,e}\equiv\frac{\int_0^d|\Psi(x)|^2dx}{J_{\rm in}}\ , \tag{5}$$

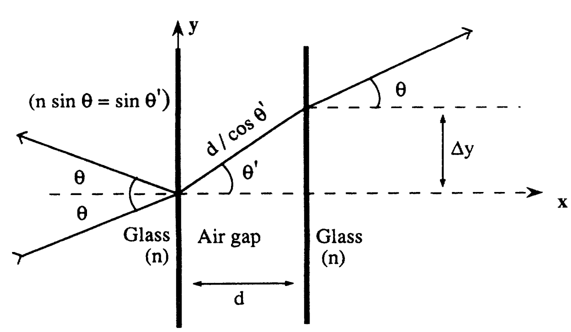

FIG. 3. In the ray optics limit for allowed traversal, the beam travels $d/\cos\theta'$ in the barrier region (where θ' is determined by Snell's law), at a velocity c.

or

$$\tau_{d,\gamma} \equiv \frac{\int_0^d \epsilon_{\rm air} |\mathcal{E}(x)|^2 dx}{J_{\rm in}} , \tag{6}$$

where $\epsilon_{\rm air} = n_{\rm air}^2$ is the dielectric constant of the barrier region, which we have taken to be one. The numerator is identical for the two problems since (1) serves to equate $\mathcal{E}$ and Ψ. The denominators will be equal, provided that we choose an effective mass such that the incident flux of the massive particle, $\hbar k_x/m$, is equal to the projection of the incident photon flux (Poynting vector) on the x axis, $cn \cos\theta$. [The physical origin of the factor of n is the n^2 in the (6)-like expression for the energy density in the first glass region, multiplied by the incident velocity c/n]. Recalling that $k_x = (n\omega/c)\cos\theta$, we find that $m = \hbar\omega/c^2$ is precisely the necessary quantity.

We should not be too surprised that this is the same relation required in the classical limit. In that limit, after all, the dwell and delay times become identical (mathematically, this is because the former can be seen as a derivative with respect to barrier height, while the latter is a derivative with respect to incident energy; for $E \gg V_0$, as well as in the WKB limit, this distinction becomes unimportant [28]). In the tunneling limit, however, the dwell may not be the same as the delay; in particular, as $E \to 0$, the former vanishes (since there is very little penetration of the particle into the barrier region) while the latter diverges.

B. The one-dimensional group delay

Now let us define the group delay for one-dimensional electron tunneling. If the barrier extends from $x=0$ to d, and the incident part of the electron wave function is

$$\psi_{k_x}(x<0) = e^{ik_x x - iEt/\hbar} , \tag{7}$$

then the field at d is $t(k_x) e^{ik_x d - iEt/\hbar}$, where $t(k_x)$ represents the transmission amplitude for the incident wave vector, and the energy $E = (\hbar k_x)^2/2m$. The total phase of the wave at $x=d$ is then

$$\begin{aligned} \phi_T &= \arg t(k_x) + k_x d - Et/\hbar \\ &\equiv \phi_d(k_x) - Et/\hbar . \end{aligned} \tag{8}$$

Consider an electron wave packet.

$$\Psi(x) = \int A(k_x) \psi k_x(x) dk_x , \tag{9}$$

with the coefficients $A(k_x)$ all real. For A sufficiently smooth and narrow band, we may use the stationary-phase approximation. The *incident* wave packet is peaked at $x=0$ at a time $t=0$, since the phase of every component vanishes there. (The wave packet as a whole has a more complicated behavior, due to the interference of the incident and reflected waves. By ignoring the latter, we are calculating the time at which the incident particle *would* reach $x=0$ if no barrier were present). By the same reasoning, the transmitted wave packet arrives at $x=d$ at a time $t=\tau_e$ such that

$$0 = \frac{\partial \phi_T}{\partial E} = \frac{\partial \phi_d(k_x)}{\partial E} - \tau_e/\hbar \tag{10}$$

implying

$$\tau_e = \hbar \frac{\partial \phi_d(k_x)}{\partial E} . \tag{11}$$

C. The two-dimensional group delay

The situation is slightly more complicated in two dimensions. Due to the translation symmetry of the system along the y axis (i.e., parallel to the interfaces), the fields have the same y dependence in all regions, and the common factor of $e^{ik_y y}$ can be dropped for the time-independent calculations, yielding equations similar to those for the electron. For calculating the group delay, however, the derivative in (10) and (11) is no longer taken with respect to E or $(\hbar k_x)^2/2m$, but rather with respect to the photon frequency ω. Since $E \neq \hbar\omega$, this will not in general yield the same value for the group delay.

On the other hand, let us analyze the *meaning* of the group delay calculated in this fashion. The wave packet described above, when multiplied by $e^{ik_y y}$, describes a pulse which is incident at the origin at $t=0$, and which arrives at $(x=d, y=0)$ at a time

$$\tau_0 = \left[\frac{\partial \phi_d}{\partial \omega} \right]_\theta . \tag{12}$$

Is this, however, the actual observable delay between the arrival of the peak at $x=0$ and its departure from $x=d$? No: the peak of the two-dimensional wave packet never crosses $(x=d, y=0)$, but rather appears at some position $(x=d, y=\Delta y)$ at a time somewhat later than τ_0. This can be seen by applying a two-dimensional version of the stationary-phase approximation. The locus of the peak is now defined by the condition that the *gradient* in k space of the phase must vanish. In the paraxial approximation, we can take the derivatives with respect to the magnitude and the direction of the k vector, i.e., with respect to ω and θ:

$$\begin{aligned} \left[\frac{\partial \phi_T}{\partial \omega} \right]_\theta &= 0 , \\ \left[\frac{\partial \phi_T}{\partial \theta} \right]_\omega &= 0 . \end{aligned} \tag{13}$$

It now becomes important to include in the total phase ϕ_T the y dependence,

$$\begin{aligned} \phi_T &= \arg t(k_x) + k_x d + k_y y - \omega t \\ &= \phi_d(k_x) + \frac{n\omega}{c} y \sin\theta - \omega t ; \end{aligned} \tag{14}$$

substituting into (13), we find for the shift Δy and the delay τ_γ of the peak

$$0 = \left[\frac{\partial \phi_d}{\partial \omega} \right]_\theta + \frac{n}{c} \Delta y \sin\theta - \tau_\gamma \tag{15}$$

and

$$0=\left[\frac{\partial\phi_d}{\partial\theta}\right]_\omega+\frac{n\omega}{c}\Delta y\cos\theta\ . \tag{16}$$

From (15), we see that there are two contributions to the delay time:

$$\tau_\gamma=\left[\frac{\partial\phi_d}{\partial\omega}\right]_\theta+\frac{n}{c}\Delta y\sin\theta\ ; \tag{17}$$

combining this with (16), we find

$$\tau_\gamma=\left[\frac{\partial\phi_d}{\partial\omega}\right]_\theta-\frac{\tan\theta}{\omega}\left[\frac{\partial\phi_d}{\partial\theta}\right]_\omega , \tag{18}$$

or

$$\tau_\gamma=\tau_0+\frac{n}{c}\Delta y\sin\theta\ . \tag{19}$$

In addition to (12)'s τ_0, the frequency derivative analogous to that of (11), there is a term which is proportional to the transverse shift.

For an undistorted wave packet, this expression has a simple physical explanation, as can be seen from Fig. 4. If at time τ_0 the intensity at $(x=d,y=0)$ reaches a maximum, we can follow the corresponding wave front as it propagates freely (at c/n) in the final medium. This wave front is perpendicular to the propagation direction, and at any given position, the maximum value of the field occurs when this wave front crosses that position. It reaches $(x=d,y=\Delta y)$ at a time $t=\tau_0+(n/c)\Delta y\sin\theta$, by inspection of the figure.

From (16), we can evaluate Δy in terms of $(\partial\phi_d/\partial\theta)_\omega$, as we have done in (18). Let us try to relate this to the electron delay time τ_e. Recalling the equivalence of the transmission amplitudes (and hence of ϕ_d) for the electron and for the photon when (1) is satisfied, and dropping the subscript, we write

$$\left[\frac{\partial\phi}{\partial\theta}\right]_\omega=\left[\frac{\partial\phi}{\partial V_0}\right]_E\left[\frac{\partial V_0}{\partial\theta}\right]_\omega+\left[\frac{\partial\phi}{\partial E}\right]_{V_0}\left[\frac{\partial E}{\partial\theta}\right]_\omega . \tag{20}$$

In order to evaluate these derivatives, we rewrite (1) as

$$V_0=(\hbar\omega)^2\frac{n^2-1}{2mc^2}\ ,$$
$$E=V_0\frac{n^2\cos^2\theta}{n^2-1}\ . \tag{21}$$

As can be seen straight away, varying the incident angle θ has no effect on the barrier height V_0, but only on the incident energy E. And it should be noted that while both V_0 and E depend on the frequency ω, their ratio does not; the essential tunneling characteristics are determined by the relation of the incident angle to the critical angle determined by n, and are not sensitive to the frequency. The first term in (20) vanishes identically, while the remaining term yields

$$\left[\frac{\partial\phi}{\partial\theta}\right]_\omega=\left[\frac{\partial\phi}{\partial E}\right]_{V_0}\left[\frac{-2n^2\cos\theta\sin\theta}{n^2-1}V_0\right]$$
$$=\left[\frac{1}{\hbar}\tau_e\right]\frac{-(n\hbar\omega)^2\cos\theta\sin\theta}{mc^2}\ . \tag{22}$$

Substituting this into (16), we find

$$\Delta y=\tau_e\frac{n\hbar\omega\sin\theta}{mc}\ . \tag{23}$$

This in turn can be substituted into (19) to yield

$$\tau_\gamma=\tau_0+\tau_e\frac{n^2\hbar\omega}{mc^2}\sin^2\theta\ , \tag{24}$$

leaving us with τ_0 to evaluate,

$$\tau_0=\left[\frac{\partial\phi}{\partial\omega}\right]_\theta=\left[\frac{\partial\phi}{\partial V_0}\right]_E\left[\frac{\partial V_0}{\partial\omega}\right]_\theta+\left[\frac{\partial\phi}{\partial E}\right]_{V_0}\left[\frac{\partial E}{\partial\omega}\right]_\theta . \tag{25}$$

As remarked above, both of these terms may contribute, since both V_0 and E are functions of ω: from (21),

$$\left[\frac{\partial V_0}{\partial\omega}\right]_\theta=\hbar^2\omega\frac{n^2-1}{mc^2}\ ,$$
$$\left[\frac{\partial E}{\partial\omega}\right]_\theta=\hbar^2\omega\frac{n^2\cos^2\theta}{mc^2}\ . \tag{26}$$

These two derivatives are of the same order of magnitude, except when $E\ll V_0$ or $E\gg V_0$. We begin by treating the former limit.

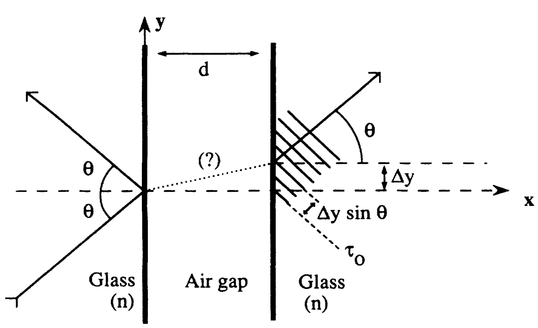

FIG. 4. Regardless of incident angle θ, the transmitted beam reaches $y=0$ earlier than it reaches $y=\Delta y$, by a delay of $(n/c)\Delta y\sin\theta$.

D. The low-energy or "deep tunneling" limit $k\ll\kappa$, or $E\ll V_0$

Near grazing incidence, $\theta\approx\pi/2$, the incident wave vector k (we henceforth drop the subscript x) is small relative to the evanescent decay constant κ; this corresponds to the limit $E\ll V_0$ for an electron. $\partial E/\partial\omega$ vanishes in this limit, removing the second term in (25), and we are left with

$$\tau_\gamma=\hbar^2\omega\frac{n^2-1}{mc^2}\left[\frac{\partial\phi}{\partial V_0}\right]_E+\hbar\omega\frac{n^2}{mc^2}\tau_e\sin^2\theta$$
$$=\hbar^2\omega\frac{n^2-1}{mc^2}\left[\frac{\partial\phi}{\partial V_0}\right]_E+\hbar^2\omega\frac{n^2}{mc^2}\left[\frac{\partial\phi}{\partial E}\right]_{V_0}\sin^2\theta\ , \quad (27)$$

where we have also used (24) and (26), and rewritten the result using (11). The prefactors are of the same order of magnitude, and $\sin\theta\to 1$; for very low energies, however, ϕ depends primarily on E and is relatively independent of V_0, allowing us to drop the first term. (The explicit expression for the phase is given in the Appendix.) Thus in the deep tunneling limit we find $\tau_0\to 0$ and

$$\tau_\gamma\approx\frac{n^2\hbar\omega}{mc^2}\tau_e\ . \quad (28)$$

This leads naturally to the observation that if an effective mass is chosen such that $mc^2=n^2\hbar\omega$, the delay time for one-dimensional tunneling of the massive particle will be the same as that for two-dimensional tunneling of the photon in the limit of grazing incidence. This agreement holds for all values of d, ranging from the thick-barrier limit $\kappa d\gg 1$ to the thin-barrier limit $\kappa d\ll 1$. In the two-dimensional problem, the delay time is due entirely to the transverse shift in this limit, as τ_0 vanishes.

E. The "critical" limit $k\gg\kappa$ or $E\approx V_0$

This limit corresponds to $\theta\approx\theta_c$, i.e., $\sin\theta\approx 1/n$. (The WKB approximation is clearly inapplicable in this region, where the "classical" traversal time diverges while the group delay remains finite.) From (26), we see that

$$\left[\frac{\partial V_0}{\partial\omega}\right]_\theta=\left[\frac{\partial E}{\partial\omega}\right]_\theta\ . \quad (29)$$

In this critical regime, the phase is primarily a function of κ and not of k, so we write

$$\left[\frac{\partial\phi}{\partial V_0}\right]_E\approx\left[\frac{\partial\phi}{\partial\kappa}\right]_k\left[\frac{\partial\kappa}{\partial V_0}\right]_E\ ,$$
$$\left[\frac{\partial\phi}{\partial E}\right]_{V_0}\approx\left[\frac{\partial\phi}{\partial\kappa}\right]_k\left[\frac{\partial\kappa}{\partial E}\right]_{V_0}\ . \quad (30)$$

However, since κ is a function of V_0-E, and not of either variable separately (that is, $\hbar^2k^2/2m\equiv V_0-E$), we see that $\partial\kappa/\partial E=-\partial\kappa/\partial V_0$, and

$$\left[\frac{\partial\phi}{\partial V_0}\right]_E=-\left[\frac{\partial\phi}{\partial E}\right]_{V_0}\ . \quad (31)$$

Combining this with (29) and substituting into (25) yields

$$\tau_0=0\ . \quad (32)$$

When we plug this into (24) and recall that $\sin\theta\approx 1/n$, we are left with the final result for the delay time in the critical regime:

$$\tau_\gamma=\tau_e\frac{\hbar\omega}{mc^2}\ , \quad (33)$$

as we saw in the classical limit in (3). This holds just above the barrier as well as just below the barrier, as can be seen by analytic continuation to imaginary values of κ. We next discuss the situation when far above the barrier.

F. The WKB or semiclassical limit

When $E>V_0$ ($\theta<\theta_c$), the wave vector in the barrier region becomes real, i.e., κ becomes imaginary. As remarked in the Introduction, the stationary-phase approximation may break down in the "allowed" regime, when multiple reflections are of comparable intensity and travel slowly in the barrier region, separating from one another at the output face. Mathematically, this is because the transmission phase develops rapid oscillations as a function of frequency [29]. The WKB or semiclassical limit corresponds to a sufficiently smoothly varying potential that multiple reflections (i.e., matching conditions at the classical turning points) are unimportant: the transmission phase depends on the local wave vector $i\kappa$ and not on k.

Even for an abrupt potential such as a square barrier, this approximation can be valid at high energies. For $E\gg V_0$, the reflection probability becomes very small, and stationary phase applies. In the corresponding optical problem, this is only possible for $n-1\ll 1$, where we may have $n^2\cos^2\theta\gg n^2-1$. In a sense, this limit is the reverse of the deep tunneling limit, in that

$$\left[\frac{\partial E}{\partial\omega}\right]_\theta\gg\left[\frac{\partial V_0}{\partial\omega}\right]_\theta\ . \quad (34)$$

However, even in the high-energy limit, we cannot be too cavalier in throwing away terms involving the right-hand side of (34) if we wish to distinguish between the equivalences which hold in deep tunneling and for the dwell time [see (28)] and those which hold in a classical, geometric approach and near criticality [see (3) and (33)]. This is because the two formulas differ only by a factor of n^2, which we have implicitly assumed to be close to 1.

As stated above, in the WKB limit, the multiple reflections are very weak; interference becomes unimportant. As can be verified explicitly from the high-energy limit of the general form of the transmission function (given in the Appendix), this means that

$$\phi\to k'd=d\sqrt{2m(E-V_0)}/\hbar\ , \quad (35)$$

where we denote the wave vector in the barrier region by $k'\equiv i\kappa$. As in the critical limit, we see that this depends only on $E-V_0$, and conclude that $(\partial\phi/\partial V_0)_E=-(\partial\phi/\partial E)_{V_0}$. In this case, however, the two terms of (25) do not cancel, because the two derivatives in (34) are no longer equal. It is readily shown that

$$\left[\frac{\partial V_0}{\partial\omega}\right]_\theta=\frac{V_0}{E}\left[\frac{\partial E}{\partial\omega}\right]_\theta$$
$$=\frac{n^2-1}{n^2\cos^2\theta}\left[\frac{\partial E}{\partial\omega}\right]_\theta\ , \quad (36)$$

where we have used (21). Using (26) and (11), we see that

$$\left[\frac{\partial\phi}{\partial E}\right]_{V_0}\left[\frac{\partial E}{\partial\omega}\right]_\theta=\tau_e\frac{n^2\hbar\omega}{mc^2}\cos^2\theta\ . \tag{37}$$

It follows that

$$\left[\frac{\partial\phi}{\partial V_0}\right]_E\left[\frac{\partial V_0}{\partial\omega}\right]_\theta=-\frac{n^2-1}{n^2\cos^2\theta}\tau_e\frac{n^2\hbar\omega}{mc^2}\cos^2\theta$$
$$=(1-n^2)\frac{\hbar\omega}{mc^2}\tau_e\ . \tag{38}$$

Substituting (37) and (38) into (24) and using the result in (25), we see that all explicit θ and n dependence cancels out, leaving us with

$$\tau_\gamma=\frac{\hbar\omega}{mc^2}\tau_e\ , \tag{39}$$

as in the critical limit. For barriers smooth enough to justify the WKB approximation, this holds throughout the regime of allowed transmission. For abrupt barriers, it holds in the high-energy limit (near-normal incidence combined with low-index contrast, in the optical case), while the stationary-phase approximation breaks down at intermediate energies due to the effects of multiple reflections.

III. TRANSMISSION AND REFLECTION TIMES FOR LOSSLESS BARRIERS

While the above discussion centered on transmission times, quite the same arguments could be made for reflection times. In fact, due to the reciprocity relations between the phase shifts for transmission and for reflection [30,31], it can be shown that for any spatially symmetric, lossless scattering system, the transmission and reflection times are *equal,* when referred to the plane of symmetry. This property follows most simply from the assumption of time-reversal invariance, although a more common treatment instead assumes unitarity. While the two assumptions are equivalent, the first is more useful for acquiring a physical understanding of this relation, and the second is more easily modified to take absorption into account. (The equality has been noted by others [13,23] for special cases such as the square barrier, and has been demonstrated elsewhere [32,33] for the general case, but seems to have escaped general attention. We rederive it here for two reasons: first, as a preliminary to the following section, in which we will generalize it to describe barriers with absorption; and second, in order to discuss the physical reason for the equality, which is typically hidden behind the admittedly straightforward manipulation of scattering matrices and phase shifts.) The assumption of spatial symmetry is valid for simple systems such as square barriers or separated prism pairs, as well as typical multilayer dielectric mirrors, which are generally terminated with high-index layers on both ends. For asymmetric systems, the reflection times for the two input directions may differ, but it can be shown that their *average* must still equal the unique transmission time. We will restrict our mathematical discussion to the symmetric case for simplicity. In this section, we will discuss the case of a lossless optical system, which is a good approximation for many dielectrics, including the glasses typically used in prisms and dielectric mirrors; the same arguments hold for real potentials in nonrelativistic quantum mechanics.

A. Phase-shift approach

This result is simplest in the one-dimensional case. Suppose a barrier has transmission and reflection amplitudes $te^{i\phi_t}$ and $re^{i\phi_r}$, respectively, where ϕ_t, ϕ_r, t, and r are taken to be real, and

$$t^2+r^2=1\ . \tag{40}$$

Further suppose that light is incident on the barrier from two directions, as in Fig. 5. In particular, let input port 1 see an incident field amplitude of $I_1=te^{-i\phi_t}$ and input port 2 see a field amplitude of $I_2=re^{-i\phi_r}$. (Physically, this corresponds to time reversing the transmitted and reflected fields which would arise had a unit-intensity field been incident on input port 2, but for now it can be considered an arbitrary choice; in the next subsection, the meaning of this choice will become clear.) The total incident power is then unity. The field exiting output port 1 is simply

$$O_1=I_1te^{i\phi_t}+I_2re^{i\phi_r}=t^2+r^2=1\ . \tag{41}$$

Since the total output power is equal to the total incident power of unity, the field exiting output port 2 must vanish. That is,

$$O_2=I_1re^{i\phi_r}+I_2te^{i\phi_t}=2rt\cos(\phi_t-\phi_r)=0\ . \tag{42}$$

Aside from the trivial case where r or t is zero (and the phase differences become meaningless quantities), this implies

$$\phi_t-\phi_r=\pm\pi/2\ . \tag{43}$$

Of course, since the group delay is the derivative of the relevant phase shift with respect to angular frequency, the difference between the transmission and reflection delays is simply

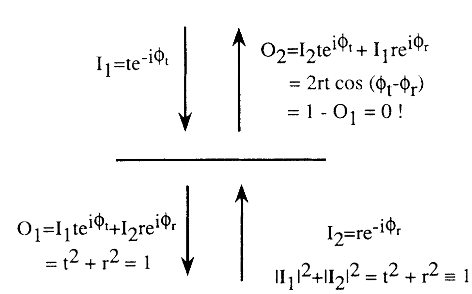

FIG. 5. By considering a special case of incident fields I_1 and I_2 on a lossless, symmetric barrier, we derive a constraint on the phases ϕ_t and ϕ_r for transmission and for reflection. The incident power is equal to 1, and this is also the power in the field O_1 which exists on the bottom; thus the field O_2 must be equal to zero.

$$\tau_t-\tau_r=\frac{d}{d\omega}(\phi_t-\phi_r)=0\ . \tag{44}$$

We have assumed that $\phi_t-\phi_r$ is continuous, a necessary assumption in order for the group delays to exist.

B. Time-reversal approach

Although the above discussion of phase shifts relied only on energy conservation, the connection with the equivalent assumption of time-reversal invariance is made apparent by the specific choice of incident fields used. The properties of the delay times follow directly from the properties of the phase shifts. Thus time-reversal invariance implies the equivalence of the transmission and delay times. Clearly, a more physical understanding of this connection should be possible by remaining in the time domain. Suppose that at time $t=0$ a packet were incident from below the barrier. Then two packets would emerge: a reflected one at time $t=\tau_r$ and a transmitted one at time $t=\tau_t$. If we time reverse this final state, we have a packet incident from below at time $t=-\tau_r$ and one incident from above at time $t=-\tau_t$. The properties of the mirror are assumed to be invariant under this transformation, so we know that the final state will involve a packet exiting from below at $t=0$, and nothing at all above. Let us consider each wave packet separately. Part of the packet incident from below is reflected, acquiring a delay of $+\tau_r$; as expected, it leaves the barrier from beneath at $t=0$. Similarly, the transmitted part of the packet incident from above leaves at $t=0$, and these two downward-going packets interfere constructively to reform the original unit-energy field. Clearly, the transmitted part of the packet incident from below and the reflected part of the packet incident from above must interfere *destructively,* as the final state does not contain any energy above the barrier. For this to be the case, they must reach the top simultaneously. The former arrives at $-\tau_r+\tau_t$ and the latter at $-\tau_t+\tau_r$; setting these two equal leads directly to the condition $\tau_t=\tau_r$. Furthermore, this argument can be trivially extended to the case of an asymmetric barrier, where τ_r may take different values, depending on the side of the barrier from which the reflection occurs (τ_t must, again by time-reversal symmetry, have a unique value). In this case, one has simply $\tau_{r1}+\tau_{r2}=2\tau_t$, i.e., that the *average* of the two reflection delay times is equal to the transmission delay time.

C. Discussion

As can be seen from consideration of Fig. 6, these arguments go through without modification for the two-dimensional case, since appropriate *total* phase functions or delay times can be defined for any incident angle, taking into account the effect of transverse shifts. (One could also follow the approach used in Sec. II of this paper, and calculate the transverse shift separately in order to determine its contribution to the total delay time. Since the transverse shift is also merely a derivative of the phase [see (16)], the above argument can be used to show that the transverse shift upon reflection is equal to the transverse shift upon transmission.)

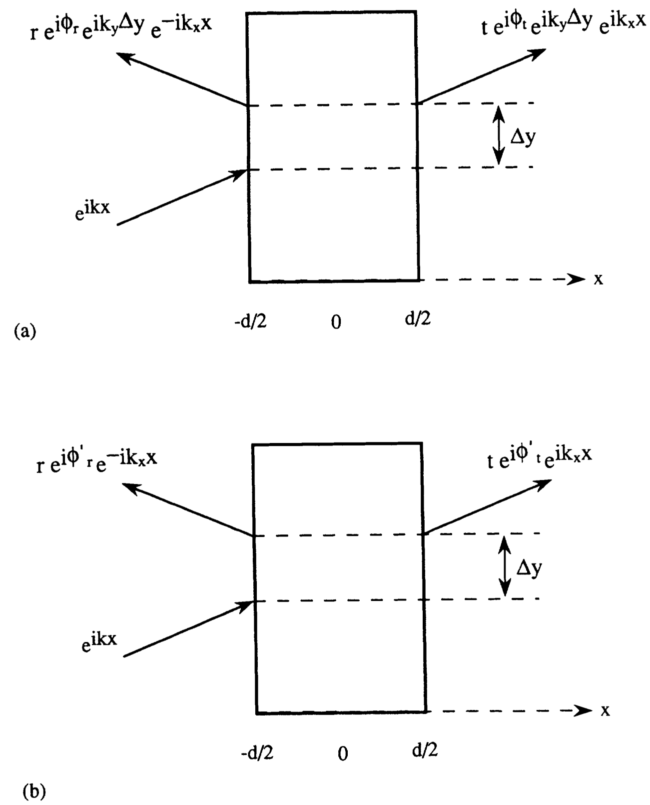

FIG. 6. (a) With the standard (one-dimensional) definition of the phase shift, the beam acquires an additional phase due to transverse displacement, as discussed in Sec. II. (b) This additional phase can be absorbed into the definition of the phase shifts; thus the discussion of Sec. III applies to two dimensions as well.

As mentioned earlier, the group delay does not grow indefinitely with the thickness of a tunnel barrier, but rather saturates at a finite limit. It is straightforward to show that the same phenomenon occurs for the transverse shift suffered by the transmitted beam. Since the transverse shift, like the temporal group delay, is the same for transmission and reflection, we see that in the limit of an infinitely thick barrier, the reflected beam is shifted transversely by a finite amount. This may be a new way to view the Goos-Hänchen effect [20]. Conversely, this equivalence can be seen as an explanation for the saturative behavior of the transmission delay. Once a barrier is more than several attenuation lengths thick, the reflection nears 100%, and it should be intuitively clear that increasing the thickness any further will have a negligible impact on the delay time for reflection; the reflected beam has not penetrated any deeper into the barrier than several attenuation lengths. The argument presented above shows that a direct consequence of this very natural asymptotic behavior is the extremely counterintuitive (but well-established) result that the delay time for transmission also becomes independent of barrier thickness in the thick-barrier limit.

It will be noted that in Fig. 3, for example, no transverse shift was indicated for the reflected beam. As expected in the ray optics limit, only the transmitted beam is shifted transversely as it crosses the gap. This should not be seen as a failure of the above relation, but merely as an example of the breakdown of the stationary-phase approximation. In the regime of classically allowed transmission, multiple reflected beams are observed, and they are typically (for wide barriers and narrow beams) separated by a distance greater than the width of the incident beam. There is no longer a unique location of the peak, and the method of stationary phase is not of use. The first spot is the product of an instantaneous Fresnel reflection, and undergoes no transverse shift or time delay.

It is important to note that, as discussed earlier, this approach has assumed a unique value of the reflection coefficient $re^{i\phi_r}$, i.e., one which does not depend on the choice of input port. For this assumption to be valid, it is necessary that the barrier be symmetric, and also that the phases (and hence the delay times) be defined in a symmetric fashion. That is, they should be referenced either to the plane of symmetry or to two planes symmetrically located with respect to it (see Fig. 7). The group delay for transmission is the delay relative to the time at which an undelayed pulse would arrive, and the group delay for reflection is the delay relative to the time at which a pulse would arrive were it instantaneously reflected upon reaching a reference plane. Clearly, the choice of this plane has no effect on the transmission delay, but may have an arbitrary effect on the value of the reflection delay. A particularly simple choice is to refer these delays not to the central plane but to the two extremities of the barrier, as in Fig. 7(b) [in the terminology of Sec. II, this corresponds to defining ϕ_t not as $\arg t$ but rather as $\phi_d \equiv \arg t + kd$; see (8)]. While not the usual definition of the transmission amplitude, this approach has the simplest interpretation in terms of delay times, in that the transmission delay becomes the total time necessary for the pulse to traverse the barrier; the reflection delay becomes the time between the arrival of the pulse at the barrier's entrance face and its departure from the same plane. These two times may seem to be physically of very different characters, but due to the symmetry of their definitions, they are actually equal. The "paradox" of superluminal group delays is only surprising for the transmitted beam, which has traversed a finite region of space. Since there is no *a priori* reason to suppose any particular finite penetration length for the reflected beam, *any* non-negative reflection delay seems plausible. While negative group delays are known to exist near absorption lines (where the reflection delay need not equal the transmission delay, due to the presence of loss), in the case of lossless scattering problems we know of superluminal group delays but no negative ones. Since anomalously short group delays are understood as a form of pulse reshaping, in which the bulk of the detected pulse originates in the early part of the incident wave, the absence of any anomalies for reflection times may be related to the fact that the transmitted part of a wave originates earlier in time than the reflected part. This suggests a connection to the Bohm–de Broglie model of quantum mechanics, in which the trajectories of different particles from a given ensemble may never cross [26,27].

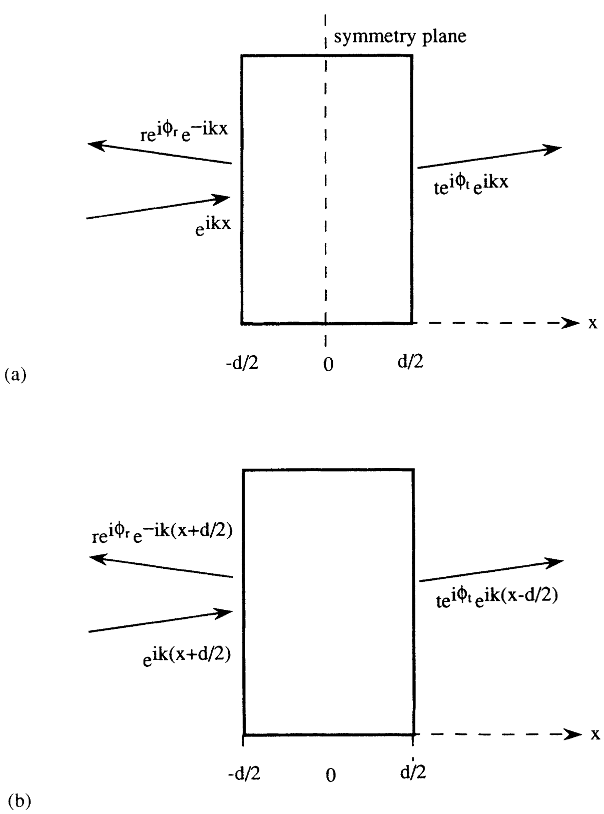

FIG. 7. (a) With the symmetry plane used as the plane of reference, the phases are defined in the familiar way, but the definition of the delays must be referred to this same plane, which is physically nonintuitive. (b) With the edges of the barrier used as the planes of reference, the phase shifts must be redefined to absorb the factor of e^{ikd}, but the delay times acquire an intuitive interpretation.

IV. TRANSMISSION AND REFLECTION TIMES WITH LOSS

The preceding section treated the case of lossless media, which is frequently a good approximation. In reality, however, the practical limit on the reflectivity of a dielectric mirror is determined by the losses due to scattering and/or absorption, however small these may be. Current mirrors can have reflectivities as high as 99.999 84% [34], in which case the transmission and the loss may be of the same order of magnitude. Since one might expect the result of the preceding section, which assumed no loss, to break down when the losses are as large as the transmitted intensity itself, we generalize the result in this section to place an upper bound on the difference between the two times, as a function of the loss. Since once energy conservation is broken even marginally, time-reversal symmetry fails categorically, we must use the phase-shift approach for this purpose.

Let us assume that we have a barrier with small but finite loss. In particular, suppose that the power loss for the case of a single incident beam is α, that is,

$$r^2+t^2=1-\alpha\leq 1 \ . \tag{45}$$

Now, it may be that the loss mechanism is completely incoherent, in the sense that *any* combination of incident beams will suffer the same reduction in total power. In that case, it is easy to show that the result from the preceding section holds exactly. Here, however, we make no such assumption. We require merely that the system is passive, i.e., that regardless of the configuration of the initial fields, the total output power is less than or equal to the total input power. Suppose now that we have fields again incident from both input ports, but now each with amplitude $1/\sqrt{2}$, such that the total incident power is again unity (it can be shown that this particular choice yields the most stringent condition on the phase shifts). In this case, the field amplitude exiting each output port is simply equal to $(r/\sqrt{2})e^{i\phi_r}+(t/\sqrt{2})e^{i\phi_t}$. The total output power is thus

$$|O_1|^2+|O_2|^2=|re^{i\phi_r}+te^{i\phi_t}|^2=r^2+t^2+2rt\cos(\phi_r-\phi_t) \ , \tag{46}$$

and this has been assumed not to exceed 1. Replacing $r^2+t^2=1-\alpha$ gives us $1-\alpha+2rt\cos(\Delta\phi)\leq 1$, or $\cos(\Delta\phi)\leq\alpha/2rt$, where we have defined the phase difference $\Delta\phi\equiv\phi_t-\phi_r$. To obtain a *lower* bound on the cosine, this same argument can be carried through for incident fields which are π out of phase rather than in phase, and we get an upper bound on the magnitude of this cosine,

$$|\cos\Delta\phi|\leq\frac{\alpha}{2rt} \ , \tag{47}$$

implying a bound on the deviation of the phase difference from $\pm\pi/2$.

Now recall that $\tau_t-\tau_r=(d/d\omega)(\phi_t-\phi_r)=(d/d\omega)\Delta\phi$. Of course, we have only an upper bound on $|\Delta\phi|$ so we cannot make any general statement about the difference between the transmission and reflection times. It is in principle possible that the phase shifts display rapid, uncorrelated oscillations within these bounds, which would allow their derivatives to differ greatly at any individual frequency. Nevertheless, we can make a statement about the average value of this difference over some bandwidth Ω. Specifically, let us define

$$\langle\tau_t-\tau_r\rangle_\Omega\equiv\frac{1}{\Omega}\int_{\omega-\Omega/2}^{\omega+\Omega/2}[\tau_t(\omega')-\tau_r(\omega')]d\omega' \ . \tag{48}$$

By substituting the definition of the group delay and performing the trivial integral, we find

$$\begin{aligned}\langle\tau_t-\tau_r\rangle_\Omega&=\frac{1}{\Omega}\int_{\omega-\Omega/2}^{\omega+\Omega/2}\frac{d}{d\omega'}(\Delta\phi)d\omega'\\&=\frac{\Delta\phi(\omega+\Omega/2)-\Delta\phi(\omega-\Omega/2)}{\Omega} \ .\end{aligned} \tag{49}$$

So long as $2rt$ remains larger than α, the cosine of the phase difference never reaches ± 1. (However, if the transmission intensity $T\equiv t^2$ *or* the reflection intensity $R\equiv r^2$ drops below approximately $\alpha^2/4$, the bound be-

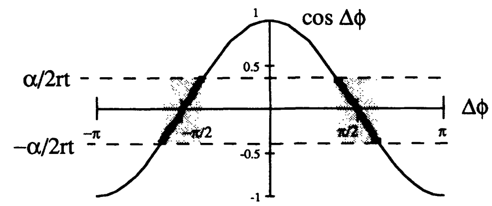

FIG. 8. The constraint $|\cos\Delta\phi|\leq(\alpha/2rt)<1$ implies that $\Delta\phi$ stays near $\pm\pi/2$ but may not move from one to the other.

comes larger than 1, and hence meaningless [35]). This restriction, coupled with the assumption of continuity, implies that $\Delta\phi$ either remains near $\pi/2$ or near $-\pi/2$, but may not move from one to the other (see Fig. 8). Then, $|\Delta\phi(\omega+\Omega/2)-\Delta\phi(\omega-\Omega/2)|\leq 2\sin^{-1}(\alpha/2rt)$. (We have assumed that $\alpha/2rt$ changes little between the limits of integration. If this is not the case, it suffices to replace it in what follows with the harmonic mean of its values at the two extremes; its value at intermediate points is irrelevant.) We therefore have

$$|\langle\tau_t-\tau_r\rangle_\Omega|\leq\frac{2}{\Omega}\sin^{-1}\frac{\alpha}{2rt} \ , \tag{50}$$

and for $\alpha\ll rt$,

$$|\langle\tau_t-\tau_r\rangle_\Omega|\leq\frac{\alpha}{rt\Omega} \ . \tag{51}$$

Thus the average of the time difference for reflection and transmission, when taken over a suitably large bandwidth, is bounded in magnitude by a time proportional to the inverse of that bandwidth. Of course, for the information to be useful, further assumptions will in general need to be made about the behavior of the delay times as a function of frequency. In particular, we have not established an upper bound on $\langle|\tau_t-\tau_r|\rangle_\Omega$, the average of the absolute value of the time difference, but only on the absolute value of the *average* time difference. Thus the possibility of the two times displaying large rapid oscillations positive and negative with respect to one another, in such a way that the *average* difference is vanishingly small, cannot be ruled out on these general grounds, although it can probably be ruled out in most specific cases.

V. CONCLUSION

In summary, we have seen that aside from the kinematic correspondence between one- and two-dimensional tunneling processes, there exists in certain limits a dynamical correspondence as well. That is, the quantum-mechanical group-delay time for a particle of mass m traversing a one-dimensional square barrier is in some limits identical to that for a photon of energy mc^2 traversing a frustrated total internal reflection barrier with index contrast n chosen according to (1). This correspondence holds both for high incident energies and for energies close to the barrier height. In the WKB limit (slowly varying potential or index of refraction), it leads to an equality between the two group delays, regardless of incident energy. It also serves to make the "dwell times" equal for the two problems regardless of incident energy or barrier width. Unlike the time-independent analogy, however, this correspondence cannot be applied universally; for very low energies, for example, the time dependence is that of a photon of energy mc^2/n^2 instead (this seems to be connected with the breakdown of the WKB or semiclassical approximation in the deep tunneling regime). Despite the differences of the two problems, these correspondences are reminiscent of the connection well known in high-energy physics between massive particles and massless particles in a higher-dimensional space. Since the two-dimensional problem can be pictured geometrically, it may be useful as an analogy for the less visualizable one-dimensional problem. The correspondence in the WKB limit does not depend explicitly on n, and may therefore be extended to arbitrary potentials so long as they are either slowly varying or small relative to the incident energy. In optics, such potentials would correspond to smoothly varying indices of refraction, or low-index-contrast multilayer dielectric structures near normal incidence, respectively. We have also demonstrated that time-reversal symmetry implies a universal equality between reflection and transmission times for lossless barriers (for photons or any other particle), and discussed the extent to which this relationship applies when loss is considered.

ACKNOWLEDGMENTS

This work was supported by the ONR under Grant Number N00014-90-J-1259. We wish to express our gratitude to Paul Kwiat for numerous discussions which have been far from fruitless.

APPENDIX: THE GROUP DELAYS IN THE OPAQUE LIMIT

In this appendix, we calculate the group delays for both the 1D Schrödinger and 2D Maxwell problems, in the "opaque" limit, defined by $\kappa d \gg 1$ and $\kappa \gg k$. (This is a strong usage of the term "opaque," restricted to low energies *and* thick barriers; either condition alone could lead to very low transmissivities.) We thus confirm the equivalence derived in this paper for one special case, and note that the delay time becomes independent of barrier thickness. We begin with the general form of the transmission amplitude, writing k for k_x for simplicity:

$$t=\frac{e^{-ikd}}{\cosh\kappa d+i[(\kappa^2-k^2)/2k\kappa]\sinh\kappa d} . \tag{A1}$$

From (8), we have

$$\begin{aligned}\phi_d(k,\kappa)&=kd+\arg t(k,\kappa)\\&=kd-kd-\tan^{-1}\left[\frac{\kappa^2-k^2}{2k\kappa}\tanh\kappa d\right] .\end{aligned} \tag{A2}$$

Assuming $\kappa d \gg 1$ at this point, $\tanh\kappa d \to 1$, and we are left with

$$\phi_d \approx \tan^{-1}\frac{k^2-\kappa^2}{2k\kappa} . \tag{A3}$$

Thus in the thick-barrier limit, ϕ_d (and therefore the group delay) becomes independent of d. When we also restrict ourselves to the case where $k \ll \kappa$, the phase reduces to

$$\phi_d \approx \tan^{-1}\frac{-\kappa}{2k} \approx -\frac{\pi}{2}+\frac{2k}{\kappa} . \tag{A4}$$

For a square barrier of height $V_0=\hbar^2k_0^2/2m$, we have $\kappa^2=k_0^2-k^2$ and thus $2\kappa d\kappa/dE=-2kdk/dE$; consequently, we can neglect the variation of κ when we take the derivative.

$$\begin{aligned}\frac{d\phi_d}{dE}&\approx\frac{\partial\phi_d}{\partial k}\frac{dk}{dE}\\&=\frac{2}{\kappa}\frac{m}{\hbar^2 k} ,\end{aligned} \tag{A5}$$

and applying (11), we find

$$\tau_e=\hbar\frac{d\phi_d}{dE}=\frac{2m}{\hbar k\kappa} . \tag{A6}$$

We now carry out the corresponding analysis for the FTIR problem. In this case, recall that the total phase at $x=d$ and $y=\Delta y$ is given by

$$\phi_T=\phi_d+\frac{n\omega}{c}\Delta y\sin\theta-\omega t , \tag{A7}$$

leading to the arrival of the peak at

$$\Delta y=-\frac{c}{n\omega\cos\theta}\frac{\partial\phi_d}{\partial\theta} . \tag{A8}$$

As remarked earlier, for the photon, the deep tunneling limit corresponds to the case of near-grazing incidence, $\theta\approx\pi/2$. Writing $\theta=\pi/2-\delta$, we have from (1) that $k_x=(\omega/c)n\cos\theta\approx(n\omega/c)\delta$ and $\kappa=(\omega/c)\sqrt{n^2\sin^2\theta-1}\approx(\omega/c)\sqrt{n^2-1}$. As before, κ remains relatively constant, and we need only concern ourselves with derivatives with respect to k. That is,

$$\begin{aligned}\Delta y&\approx-\frac{c}{n\omega\cos\theta}\frac{\partial\phi_d}{\partial k_x}\frac{dk_x}{d\theta}\\&=-\frac{c}{n\omega\cos\theta}\frac{2}{\kappa}\frac{-n\omega}{c}\\&=\frac{2}{\kappa\cos\theta} .\end{aligned} \tag{A9}$$

The group delay

$$\begin{aligned}\tau_\gamma&=\frac{\partial\phi_d}{\partial\omega}+\frac{n}{c}\Delta y\sin\theta\\&\approx\frac{2n\delta}{c\kappa}+\frac{2n}{c\kappa\cos\theta}\\&\approx\frac{2n^2\omega}{c\kappa ck} .\end{aligned} \tag{A10}$$

Notice that at grazing incidence, this time is entirely due to the transverse shift; the time it takes the field at $y=0$ to reach a maximum is negligible compared to the additional time the peak takes to reach Δy. (As shown in Sec. II, it is smaller than τ_γ by the simple factor of $\cos^2\theta$, here δ^2.) If we now substitute $\omega=m_{\mathrm{eff}}c^2/\hbar n^2$, we find

$$\tau_\gamma=\frac{2m_{\mathrm{eff}}}{\hbar k\kappa} , \tag{A11}$$

exactly as for the electron of (A6).

[1] M. Büttiker and R. Landauer, Phys. Rev. Lett. **49**, 1739 (1982).
[2] E. H. Hauge and J. A. Støvneng, Rev. Mod. Phys. **61**, 917 (1989), and references therein.
[3] R. Landauer, Nature (London) **341**, 567 (1989).
[4] H. A. Fertig, Phys. Rev. Lett. **65**, 2321 (1990).
[5] M. Büttiker, in *Electronic Properties of Multilayers and Low Dimensional Semiconductors,* edited by J. M. Chamberlain, L. Eaves, and J. C. Portal (Plenum, New York, 1990), p. 297.
[6] R. Landauer, Ber. Bunsenges. Phys. Chem. **95**, 404 (1991).
[7] E. H. Hauge, J. P. Falck, and T. A. Fjeldly, Phys. Rev. B **36**, 4203 (1987).
[8] W. Jaworski and D. M. Wardlaw, Phys. Rev. A **37**, 2843 (1988).
[9] C. R. Leavens and G. C. Aers, Phys. Rev. B **39**, 1202 (1989).
[10] A. Ranfagni, D. Mugnai, P. Fabeni, G. P. Pazzi, G. Naletto, and C. Sozzi, Physica B **175**, 283 (1991), and references therein.
[11] A. Enders and G. Nimtz, J. Phys. (Paris) I **3**, 1089 (1993).
[12] R. Y. Chiao, P. G. Kwiat, and A. M. Steinberg, Physica B **175**, 257 (1991).
[13] Th. Martin and R. Landauer, Phys. Rev. A **45**, 2611 (1992).
[14] A. M. Steinberg, P. G. Kwiat, and R. Y. Chiao, Phys. Rev. Lett. **71**, 708 (1993).
[15] T. K. Gaylord, G. N. Henderson, and E. N. Glytsis, J. Opt. Soc. Am. B **10**, 333 (1993).
[16] As has been discussed in many papers [7–9], a "local" delay time of this sort cannot be directly measured, but must be extrapolated from observations of the wave packet far from the barrier. As the behavior of the packet outside the barrier region may be a function of energy, and the transmitted packet may have a different spectrum than the incident one, this leads to considerable subtleties. Here, however, we leave aside these issues in order to discuss the amount of *extra* time a *given* frequency component takes to reach an arbitrary point on the far side of the barrier, compared to the time it would have taken had the particle instantaneously "hopped" from the entrance face of the barrier to its exit face. Similarly, we make no attempt to identify this delay time with the "dwell" in any particular region of space, except in the classical limit where this becomes relatively straightforward. Therefore we leave aside the issue of the complicated self-interference of the incident and reflected wave packet—which has been described as introducing some delay before the particle even reaches the barrier—using as our reference the time at which the incident peak *alone* would arrive at the position in question.
[17] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), p. 302.
[18] M. Büttiker and R. Landauer, Phys. Scr. **32**, 429 (1985).
[19] See J. Opt. Soc. Am. B **10** (2) (1993), special issue on photonic band gaps, edited by C. M. Bowden, J. P. Dowling, and H. O. Everitt.
[20] A. K. Ghatak, *Contemporary Optics* (Plenum, New York, 1978), Chap. 11, pp. 331–352.
[21] C. K. Carniglia and L. Mandel, Phys. Rev. D **3**, 280 (1971).
[22] C. K. Carniglia and L. Mandel, J. Opt. Soc. Am. **61**, 1035 (1971).
[23] A. Ghatak and S. Banerjee, Appl. Opt. **28**, 1960 (1989).
[24] S. Bosanac, Phys. Rev. A **28**, 577 (1983).
[25] J. Spector, H. L. Stormer, K. W. Baldwin, L. N. Pfeiffer, and K. W. West, Appl. Phys. Lett. **56**, 2433 (1990).
[26] A. M. Steinberg, P. G. Kwiat, and R.Y. Chiao, Found. Phys. Lett. (to be published).
[27] C. R. Leavens, Solid State Commun. **76**, 253 (1990), and references therein.
[28] M. Büttiker, Phys. Rev. B **27**, 6178 (1983).
[29] In the tunneling regime, by contrast, the imaginary value of the k vector turns this oscillation into a hyperbolic function. The phase varies smoothly from $\pi/2$ at $E=V_0$ to $-\pi/2$ at $E=0$, and the stationary-phase approximation does not break down.
[30] C. Cohen-Tannoudji, B. Diu, and F. Lalöe, *Quantum Mechanics* (Wiley, New York, 1977), p. 363.
[31] Z. Y. Ou and L. Mandel, Am. J. Phys. **57**, 66 (1988).
[32] J. P. Falck and E. H. Hauge, Phys. Rev. B **38**, 3287 (1988).
[33] M. Ya. Azbel, Phys. Rev. B **28**, 4106 (1983).
[34] G. Rempe, R. J. Thompson, H. J. Kimble, and R. Lalezari, Opt. Lett. **17**, 363 (1992).
[35] For low loss, this will typically only occur in regions near where r vanishes, making the reflection phase shift (and delay time) undefined; t can never vanish.

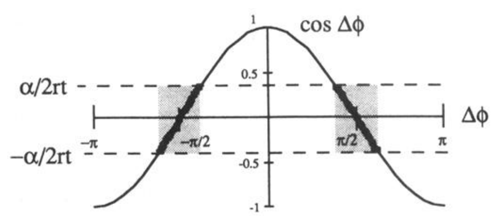

FIG. 8. The constraint $|\cos\Delta\phi| \leq (\alpha/2rt) < 1$ implies that $\Delta\phi$ stays near $\pm\pi/2$ but may not move from one to the other.