Physical Mechanism for Apparent Superluminality in Barrier Tunneling

Herbert G. Winful, *Fellow, IEEE*

***Abstract*—Tunneling without distortion is not a propagation phenomenon but a quasistatic process in which the slowly varying envelope of the input pulse modulates the amplitude of a standing wave in the barrier. Because the envelope has a bandwidth that is small compared to the cutoff frequency of the barrier it does not propagate; it merely stands and waves. The envelope of the reflected and transmitted fields can adiabatically follow the envelope of the input pulse with only a small delay proportional to the energy stored in the barrier. The Hartman effect, in which the delay becomes independent of barrier length, is explained on the basis of saturation of stored energy with increasing length. Because the delay is not a propagation delay, it should not be linked to a velocity of propagation. True evanescent waves have zero speed and not infinite velocity as often asserted.**

***Index Terms*—Barrier tunneling, energy storage, evanescent waves, modulated standing waves, photonic bandgap structures, superluminality.**

I. Introduction

THERE DOES not appear to be universal agreement on the physical mechanism responsible for the apparent superluminal tunneling of wave packets through a forbidden region [1], [2]. Indeed, 70 years after MacColl suggested that barrier traversal takes no appreciable time [3], the nature of the tunneling process is still considered "a poorly resolved mystery" [4]. Among the proposed mechanisms are pulse reshaping and interference effects which cause the front of the pulse to be attenuated less than the rear, leading to a forward shift of the center of gravity of the entire pulse [2], [5]–[7]. According to this theory, the entire transmitted pulse is reconstructed from just the weak leading edge of the much larger input pulse, with the barrier in effect performing an analytic continuation [7]. While the reshaping description may account for apparent superluminality in situations where the pulse is indeed reshaped, it does not explain the distortionless tunneling that has been observed in many experiments [5], [8]–[10]. It also does not explain, in any simple physical way, some paradoxical phenomena such as the "Hartman effect" in which the tunneling time becomes independent of barrier length [8], [11], [12]. This effect has led to a number of claims that evanescent waves travel with infinite velocity and thus tunnel in zero time [12], [13]. The time scale of the tunneling process still remains a matter for debate and the literature is replete with assorted tunneling time definitions [14].

Manuscript received October 31, 2002; revised December 5, 2002. This work was supported in part by the National Science Foundation under Grants PHY99-07949 and 0114336.

The author is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: winful@eecs.umich.edu).

Digital Object Identifier 10.1109/JSTQE.2002.807972

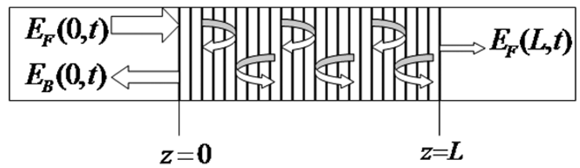

Fig. 1. Schematic of a photonic bandgap structure.

In recent papers, we have presented a physical mechanism for the tunneling process that resolves much of the mystery surrounding it [15], [16]. Here, we present a more detailed account. The key to unraveling the mystery is the realization that barrier tunneling is not a propagation phenomenon. It is a *quasistatic* process in which the envelope of the incident pulse modulates the amplitude of an exponentially attenuating standing wave [15]. Because it is a standing wave, the field at all points in the barrier can move up and down in phase with the incident envelope or with a small phase lag related to energy storage in the barrier. We find that the entire input pulse contributes to the transmitted pulse as it modulates the stored energy in the barrier. Because this is a quasistatic process involving standing waves, one cannot speak of evanescent waves as propagating with infinite speed. As a matter of fact, they do not propagate at all. The modulated-standing-wave picture of barrier tunneling makes it possible to resolve all the outstanding paradoxes in the tunneling process. Because this subject is controversial and has been debated for years, our presentation will be somewhat pedantic as we seek to clarify a rather murky subject.

II. The Barrier Model

Tunneling experiments in one spatial dimension have mostly been carried out on photonic bandgap structures (PBGs) excited within their stop band [5], [8]–[10] or on waveguides operating below cutoff [12], [17]. The waveguide tunneling problem has been analyzed using the Klein–Gordon equation for the electric field in the guide [18]–[20]. It turns out that the same equation describes the envelopes of forward and backward waves in a periodic dielectric structure, also known as a PBG. We will thus refer specifically to a PBG in the derivations while keeping in mind the general applicability of the resulting equations. Fig. 1 shows a one-dimensional PBG with a periodic variation in the refractive index given by

$$n = n_0 + n_1 \cos(2\beta_0 z) \tag{1}$$

where n_0 is the unperturbed refractive index, $n_1 \ll n_0$ is the amplitude of the index perturbation, $\beta_0 = n_0\omega_0/c$ is the Bragg wavenumber, and ω_0 is the carrier angular frequency that satisfies the Bragg condition for the structure. The periodic structure extends from $z = 0$ to $z = L$ and is embedded in a homogeneous region of refractive index n_0. The unperturbed medium is assumed lossless and dispersionless. This neglect of material dispersion is appropriate since the structural dispersion due to the periodic perturbation is orders of magnitude larger than that of the material [21].

The periodic index perturbation scatters forward waves into backward waves, hence we take the complex electric and magnetic fields within the structure as

$$E(z,t) = E_F(z,t)e^{i(\beta_0 z-\omega_0 t)} + E_B(z,t)e^{-i(\beta_0 z+\omega_0 t)} \quad (2)$$

$$H(z,t) = \left(\frac{1}{\eta}\right)\left[E_F(z,t)e^{i(\beta_0 z-\omega_0 t)} - E_B(z,t)e^{-i(\beta_0 z+\omega_0 t)}\right] \quad (3)$$

where E_F and E_B are the forward and backward components of the field envelopes, and $\eta = \sqrt{\mu/\varepsilon}$ is the intrinsic impedance of the unperturbed medium. The physical fields are the real parts of the above. Within the slowly varying envelope approximation, use of (1)–(3) in Maxwell's equations leads to the following coupled-mode equations for the forward and backward fields [21]:

$$\frac{\partial E_F}{\partial z} + \frac{1}{v}\frac{\partial E_F}{\partial t} = i\kappa E_B \quad (4a)$$

$$\frac{\partial E_B}{\partial z} - \frac{1}{v}\frac{\partial E_B}{\partial t} = -i\kappa E_F. \quad (4b)$$

Here, $\kappa = \pi n_1/\lambda_0$ describes the coupling between forward and backward waves and $v = c/n_0$ is the group and phase velocity in the unperturbed medium. It should be noted that the carrier waves $\exp[i(\beta_0 z \pm \omega_0 t)]$ are solutions of the homogeneous wave equation for the lossless unperturbed medium and are always propagating modes. The envelope functions modulate the amplitude of these carrier waves and determine how much of the carrier will be transmitted by the barrier. Because these envelope functions are simultaneously present within the medium and are coupled, they cannot be accessed separately. In fact, this decomposition is not unique and so one should not take the expressions "forward wave" and "backward wave" too literally. The physical measurable fields inside the medium are composed of sums and differences of these functions, as given in (2) and (3). Outside the periodic structure, the forward and backward envelopes are uncoupled and hence can be associated with incident, transmitted, and reflected waves.

Contact with other tunneling structures may be made by noticing that (4a) and (4b) combine to yield a Klein–Gordon equation (KGE) for the forward envelope

$$\frac{\partial^2 E_F}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2 E_F}{\partial t^2} = \kappa^2 E_F. \quad (5)$$

A similar equation holds for the backward modulation E_B. The KGE describes many wave phenomena that involve a cutoff frequency below which wave propagation does not occur. These include waves in torsionally coupled pendula, electromagnetic waves in plasmas and waveguides, and relativistic matter waves [22]. The model used here therefore applies to tunneling phenomena in general. For baseband signals without a carrier, we set $\omega_0 = 0$ and then the KGE applies to the complete electric field and not an envelope.

Since (5) describes the evolution of a slow modulation riding on top of a carrier, one may ask how the bandwidth of this modulation affects the nature of wave propagation within the structure. To that end, we consider a traveling wave modulation of frequency $\Omega \ll \omega_0$ and wavenumber $K \ll \beta_0$. Inserting $E_F - \exp[i(Kz - \Omega t)]$ into the KGE we obtain the dispersion relation

$$K^2 = \frac{\Omega^2}{v^2} - \kappa^2. \quad (6)$$

Any wave phenomenon described by such a dispersion relation leads to the KGE through the operator transformation $K \rightarrow i\partial/\partial z, \Omega \rightarrow i\partial/\partial t$. The cutoff angular frequency for the system is defined as

$$\Omega_c = \kappa v. \quad (7)$$

If the modulation frequency exceeds Ω_c, the wavenumber K is real and the envelope propagates through the structure.

On the other hand, for a slow modulation such that $\Omega < \Omega_c$, the wavenumber becomes imaginary and the wave is evanescent; the envelope does not propagate but merely decays exponentially with distance as

$$\exp(\pm\gamma z)\exp(-i\Omega t) \quad (8)$$

where the attenuation constant is given by

$$\gamma = \sqrt{\frac{(\Omega_c^2 - \Omega^2)}{v^2}}. \quad (9)$$

In (8) the plus sign governs an envelope that attenuates in the negative z direction, while the minus sign describes an envelope that attenuates in the positive z direction. We emphasize, however, that (8) does not describe wave propagation but rather a process in which the field at every point in z moves up and down in phase. This is exactly the behavior of a standing wave. Thus, a slow modulation results in a standing wave with an exponentially decaying envelope. True evanescent waves therefore do not propagate. They merely stand and wave [22].

III. Phase Lag and Group Delay

The coupled mode equations immediately tell us that the backward wave will have a phase lag and a delay with respect to the forward wave. Consider a forward evanescent mode $E_F = \exp(-\gamma z)\exp(-i\Omega t)$. This is a mode of an infinitely long barrier. From (4a) the backward wave is given by

$$E_B = \frac{1}{i\kappa}\left[\frac{\partial E_F}{\partial z} + \frac{1}{v}\frac{\partial E_F}{\partial t}\right] = \frac{1}{i\kappa}\left[-\gamma - \frac{i\Omega}{v}\right]e^{-\gamma z}e^{-i\Omega t}. \quad (10)$$

The phase of the backward wave with respect to the forward wave is therefore

$$\phi_0 = \frac{\pi}{2} + \tan^{-1}\left(\frac{\Omega}{\gamma v}\right). \quad (11)$$

It is the phase of the local reflection coefficient induced by the coupling κ. It is not a propagation phase shift since the envelope does not propagate. This phase shift corresponds to a time shift in the backward evanescent mode of ϕ_0/Ω. The peak of a narrowband pulse reflected from an infinite barrier will suffer a group delay of

$$\tau_g = \left.\frac{d\phi_0}{d\Omega}\right|_{\Omega=0} = \left.\frac{1}{\gamma v}\right|_{\Omega=0} = \frac{1}{\kappa v} \tag{12}$$

assuming no index mismatch at the interface. This delay is not a propagation delay. We will see that it is related to energy storage.

In general, if the forward standing wave in an infinitely long barrier consists of an exponentially decaying spatial part multiplying a slow temporal function, i.e., $E_F = e^{-\kappa z}F(t)$ then, again from (4a), the backward wave is given by

$$\begin{aligned} E_B(z,t) =& \frac{1}{i\kappa}\left[-\kappa E_F(z,t) + \frac{1}{v}\frac{\partial E_F(z,t)}{\partial t}\right] \\ \approx& iE_F\left(z, t - \frac{1}{\kappa v}\right). \end{aligned} \tag{13}$$

Thus, the coupling gives rise to a backward wave that is everywhere in phase quadrature to the forward wave and lags behind it by a delay of $1/\kappa v$, independent of length. The delay depends only on the "shape" of the mode and is equal to the inverse of the cutoff angular frequency.

IV. Reactive Nature of Coupling

The coupling seen here is a reactive, purely conservative form of coupling which leads to energy storage as opposed to dissipation. A consequence of this energy storage is a phase shift and delay of the backward wave with respect to the forward wave.

The coupled-mode equations lead to the following conservation laws:

$$\frac{\partial}{\partial z}\left[v\left(|E_F|^2 - |E_B|^2\right)\right] = -\frac{\partial}{\partial t}\left[|E_F|^2 + |E_B|^2\right] \tag{14}$$

$$\begin{aligned} &\frac{\partial}{\partial z}\left[v\left(|E_F|^2 + |E_B|^2\right)\right] + \frac{\partial}{\partial t}\left[|E_F|^2 - |E_B|^2\right] \\ &\quad = -4\kappa v \mathrm{Im}(E_F^* E_B). \end{aligned} \tag{15}$$

These are obtained by multiplying (4a) and (4b) by the complex conjugates of E_F and E_B, respectively, and then forming sums and differences. Equation (14) is simply Poynting's theorem in differential form, relating the time rate of decrease of energy density at a point to the divergence of the power flux at that point. Power flow is conserved in this form of coupling. In the second equation, the term on the right-hand side is proportional to the difference between stored magnetic and electric energies. This term accounts for reactive energy storage and shows that the spatial decrease in total stored energy is related not just to real power flow but to a circulating contribution which is a reactive power. Since the group delay for the infinite barrier is $\tau_g = 1/\kappa v$, expressing the right-hand side of (14) as $-4\mathrm{Im}(E_B E_F^*)/\tau_g$ suggests that this reactive energy is stored

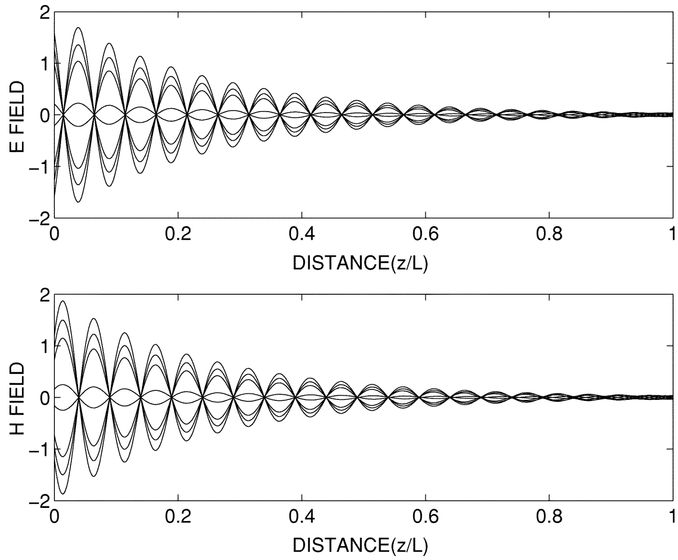

Fig. 2. Electric and magnetic fields in a finite barrier over several optical cycles of the optical carrier wave. Here $\beta_0 L = 20\pi$ and $\kappa L = 4$.

and released on a time scale given by the group delay. This connection is placed on firmer ground in the Appendix. This term also depends on the phase difference between forward and backward components. As shown in Section X, the phase difference is related to the stored energy and, hence, the group delay.

The reactive nature of the coupling can also be seen by considering the complex Poynting vector, defined for time-harmonic fields as

$$\begin{aligned} S =& \frac{1}{2}EH^* \\ =& \frac{1}{2\eta}\left[\left(|E_F|^2 - |E_B|^2\right) - 2i\mathrm{Im}\left(E_F E_B^* e^{i2\beta z}\right)\right]. \end{aligned} \tag{16}$$

The real part of the complex Poynting vector gives the time-averaged power flow per unit area, while the imaginary part accounts for reactive power which is stored and pulsates between electric and magnetic forms. It is proportional to the difference between the stored magnetic and electric energy densities. The term on the right-hand side in (15) thus describes the envelope of the reactive power flow and shows that it is indeed the coupling that is responsible for this reactive power. The reactive power is maximized when the forward and backward envelopes are 90° out of phase. For the infinite barrier, where $|E_F|^2 = |E_B|^2$ on a time-averaged basis, the power is purely reactive. What we have inside the barrier is a complete standing wave with the electric field and magnetic field distributions shifted in space as well as in time because of the phase difference. If the barrier is finite, the standing wave is nearly complete in the front half of the barrier where the forward and backward intensities are nearly equal. At the exit, the backward wave is zero and so the forward carrier wave can exist as a traveling wave in that region, transporting real power. Fig. 2 shows the standing wave profiles for an evanescent wave over several optical cycles as computed from exact solutions of (4) [15]. In the tunneling process, the slow envelope of the incident pulse modulates the amplitude of this standing wave.

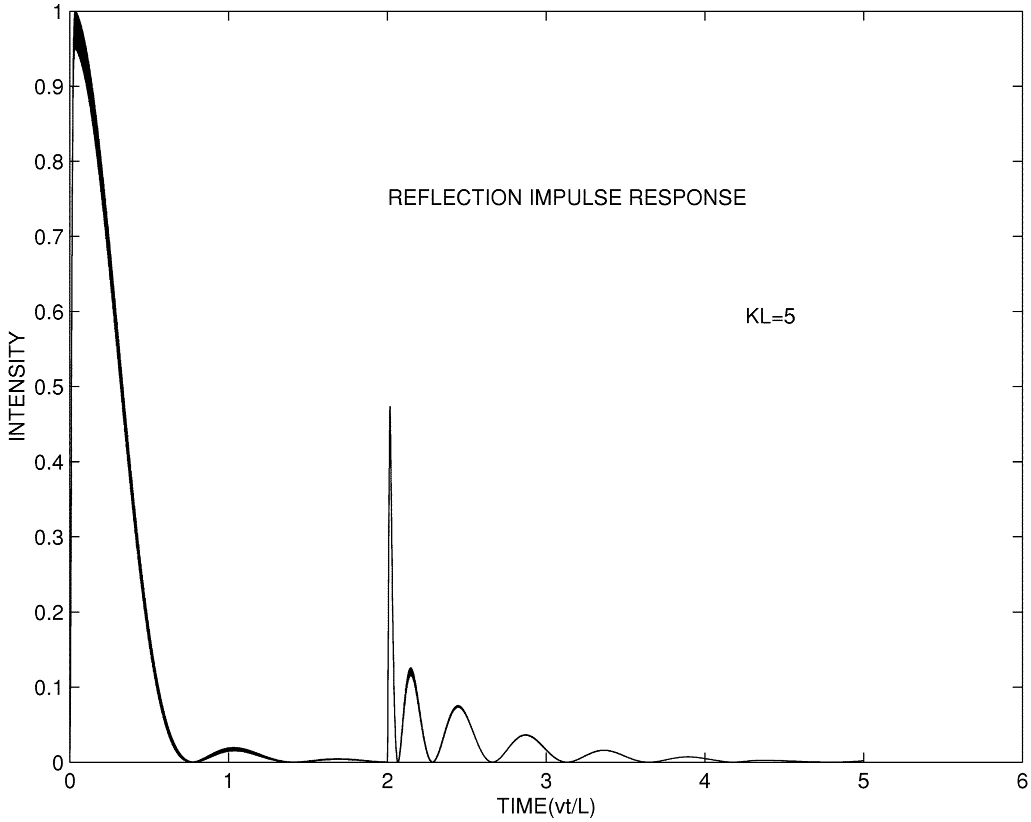

Fig. 3. Reflection impulse response for a finite length periodic structure.

V. Impulse Response

The KGE has the causal Green's function [23]

$$g(z,t)=\delta\left(t-\frac{z}{v}\right)-\frac{\Omega_c z J_1\left(\Omega_c\sqrt{t^2-\left(\frac{z}{v}\right)^2}\right)}{v\sqrt{t^2-\left(\frac{z}{v}\right)^2}}\Theta\left(t-\frac{z}{v}\right). \tag{17}$$

Here, $J_1(x)$ is a first-order Bessel function and $\Theta(t)$ is the unit step function. This is the oscillatory impulse response for an infinitely long barrier. It is a good approximation for the impulse response of a reasonably opaque barrier, an exact form of which can be found in [19], for example. Because it is perfectly causal, true superluminality is precluded. Any superluminality seen here will have to be of the "apparent" variety.

With use of the asymptotic expansion of the Bessel function we find that for large times, the oscillations of the impulse response occur roughly at the cutoff frequency $f_c=\Omega_c/2\pi$. Fig. 3 shows the intensity reflection impulse response for a finite barrier as computed using the coupled-mode equations. Note the spike at one round trip which corresponds to the discontinuity at the exit experienced by the total field. The reflectivity impulse response is a useful tool in the characterization of fiber Bragg gratings [24]. Since the cutoff frequency is related to the coupling constant through $\Omega_c=\kappa v$, a measurement of the zeroes of the impulse response can be used to determine the coupling constant. It is important to note that whenever a barrier is excited with a pulse whose bandwidth exceeds the cutoff frequency, the output will "ring" at the cutoff frequency. In (3), the intensity is plotted and not the electric field, hence the oscillations are at twice the cutoff frequency.

VI. Step Response

The step response of the infinite barrier can be readily obtained by integrating the impulse response. Here, we obtain the step response of a finite barrier by solving the coupled mode equations numerically so we can follow the evolution of a step through the barrier. By this means we see how a front propagates and how fields behind the front are rejected by the barrier. Fig. 4(a) shows that the front propagates at c since it is dominated by the high-frequency components. Behind the front is light that has undergone various amounts of scattering. The energy density in the barrier near the entrance quickly builds up behind the front in a characteristic near-exponential profile. The energy is thus stored at the front end. This exponential distribution that characterizes the steady-state field is approached within one transit time. Fig. 4(b) shows the transmitted and reflected intensities. At the output the front is transformed into a sharp spike with some trailing ringing at the cutoff frequency as the barrier acts as a differentiator or high-pass filter. The reflected intensity at the input exhibits overshoot and ringing in its approach to steady state.

VII. Response to a Sinusoidal Modulation

Now consider a sinusoidal envelope modulation of the form $E_F(0,t)=E_0\cos\Omega t$, with $\Omega<\Omega_c$, and impose the condition $E_B(L,t)=0$. The modulation is turned on at time $t=0$.

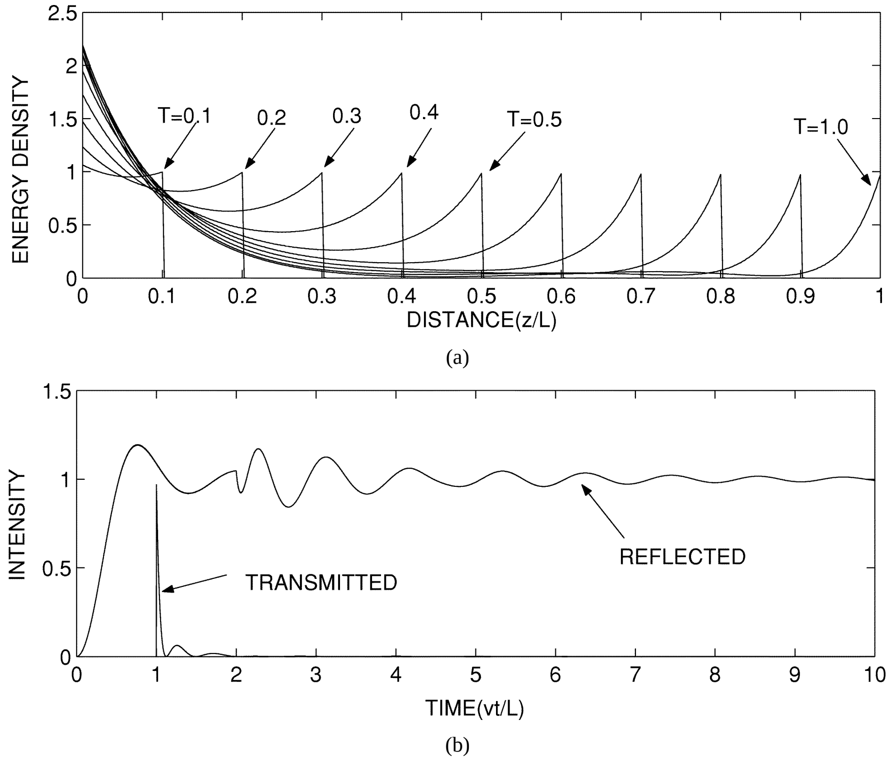

Fig. 4. (a) Propagation of the front of a step. (b) Step response for reflected and transmitted intensities.

Fig. 5 shows the transmitted sinusoidal envelope modulation as obtained from a numerical solution of the coupled-mode equations. The vertical dashed lines indicate the locations of the maxima of the squared input modulation. After an initial transient, the output reaches a sinusoidal steady state with the peaks of the reflected and transmitted sinusoidal modulations time shifted with respect to the input. This is actually the basis for a practical method of characterizing group delay in fibers that involves detecting the phase shift suffered by a modulated cw beam [25], [26]. In this simulation, for $\kappa L = 4$, we find a time shift (normalized by the transit time L/v) of approximately $1/\kappa L = 0.25$, which is the limiting group delay of (12).

Analytical solutions of the coupled-mode equations in the sinusoidal steady state are of the form

$$\begin{Bmatrix} E_F(z,t) \\ E_B(z,t) \end{Bmatrix} = \mathrm{Re}\begin{Bmatrix} E_F(z) \\ E_B(z) \end{Bmatrix} e^{-i\Omega t}. \tag{18}$$

The spatial distributions of the envelope functions are found to be

$$E_F(z) = E_0 \frac{\gamma}{g}\left[\cosh\gamma(z-L) + i\frac{\Omega}{\gamma v}\sinh\gamma(z-L)\right] \tag{19a}$$

$$E_B(z) = -i\frac{[E_0\kappa\sinh\gamma(z-L)]}{g} \tag{19b}$$

where $g = \gamma\cosh\gamma L - i(\Omega/v)\sinh\gamma L$. The barrier amplitude transmission coefficient is

$$T = \frac{E_F(L)}{E_0} = \left(\frac{\gamma}{|g|}\right)e^{i\phi_t} \tag{20}$$

the phase of which is given by

$$\phi_t = \tan^{-1}\left[\left(\frac{\Omega}{\gamma v}\right)\tanh\gamma L\right]. \tag{21}$$

The amplitude reflection coefficient is

$$R = \left(\frac{\kappa\sinh\gamma L}{|g|}\right)e^{i\phi_r} \tag{22}$$

where

$$\phi_r = \tan^{-1}\left[\left(\frac{\Omega}{\gamma v}\right)\tanh\gamma L\right] + \frac{\pi}{2}. \tag{23}$$

To see how a sinusoidal modulation tunnels through the barrier, we write out the solutions in (19) explicitly as functions of space and time by inserting in (18)

$$\begin{aligned} E_F(z,t) =& \frac{E_0\cos\phi_t}{\cosh\gamma L} \\ &\times\{\cosh\gamma(z-L)\cos(\Omega t-\phi_t) \\ &+\frac{\Omega}{\gamma v}\sinh\gamma(z-L)\sin(\Omega t-\phi_t)\Big\} \end{aligned} \tag{24a}$$

$$\begin{aligned} E_B(z,t) =& -\left[\frac{\kappa E_0\cos\phi_t}{\gamma\cosh\gamma L}\right] \\ &\times\sinh\gamma(z-L)\sin(\Omega t-\phi_t). \end{aligned} \tag{24b}$$

These are envelope standing waves. The cosh and sinh functions are normal modes of the finite barrier and as a result they move up and down in their entirety in response to the input modulation. They do not go anywhere. Fig. 6 shows the evolution of the "forward" envelope at successive instants of time between

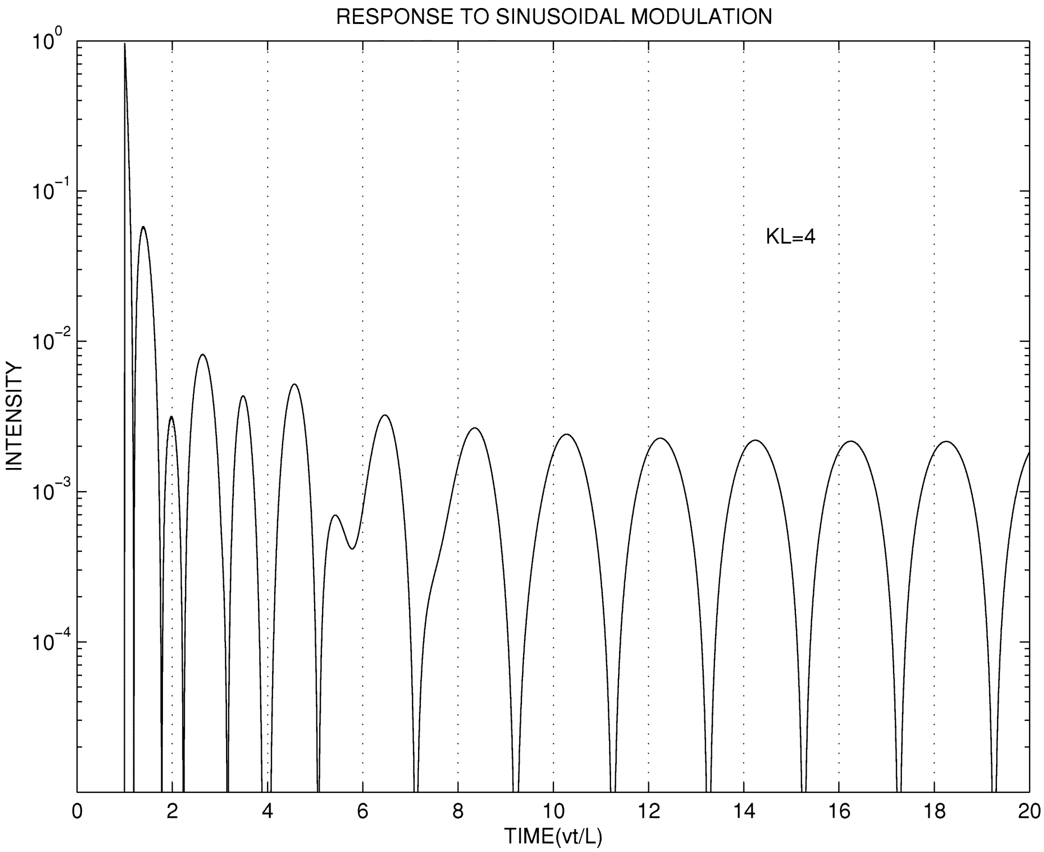

Fig. 5. Output corresponding to a sinusoidal modulation of input envelope.

$t = 11$ and $t = 13$. The input modulation has a peak at $t = 12$, as seen in Fig. 5. The entire envelope is seen to rise and fall with the input modulation. Clearly, this is not a propagation phenomenon. The total envelope field thus consists of two standing wave modes in quadrature with each other. At each point in the barrier, the temporal variation of the field is a superposition of independent contributions from each mode. The beating between these two modes allows for energy stored in one location to be released downstream. Each mode has an additional phase shift of ϕ_t with respect to the driving modulation. Since the envelope does not propagate in the sinusoidal steady state, the phase shifts seen here are not due to propagation but are a result of the reactive nature of the coupling, which leads to energy storage, and of any impedance mismatch at the boundaries of the barrier. As a result of these phase shifts the peaks of the forward and backward envelope modulations are delayed in time with respect to the driving modulation. This delay is the group delay or envelope delay (also referred to as the phase time) and is given by $d\phi_t/d\Omega$ for the transmitted wave and by $d\phi_r/d\Omega$ for the reflected wave. For the transmitted $(E_F(L,t))$ and reflected $(E_B(0,t))$ fields we find the same group delay

$$\tau_g = \frac{1}{v}\left[\frac{\kappa^2}{\gamma^2}\frac{\tanh\gamma L}{\gamma} - L\left(\frac{\Omega}{\gamma v}\right)^2 \mathrm{sech}^2\gamma L\right]\cos^2\phi_t \quad (25)$$

where

$$\cos^2\phi_t = \frac{1}{1+\left(\frac{\Omega}{\gamma v}\right)^2\tanh^2\gamma L}.$$

Fig. 7 shows the intensity transmission function $|T(\Omega)|^2$ and the group delay normalized by the transit time $t_{tr} = L/v$. Luminal delays correspond to $\tau_g/T_{tr} = 1$. Superluminal delays $(\tau_g/T_{tr} < 1)$ occur for normalized detunings within the stopband. We emphasize that this superluminal delay is not a propagation delay since the envelope does not propagate when the modulation frequency is below cutoff. The behavior here is similar to that of a simple RC (resistance-capacitance) high-pass filter. The barrier is acting as a lumped capacitor with coupling to the outside world providing an effective dissipation.

VIII. Quasistatics of Barrier Tunneling

A barrier such as a PBG structure or an undersized waveguide acts as a high-pass filter with a transfer function

$$T(\Omega) = |T(\Omega)|\, e^{i\phi_t(\Omega)}. \quad (26)$$

Such a filter will transmit a pulse with a delay $\tau_g = d\phi_t/d\Omega$ but without distortion if the magnitude of the transfer function is constant and its phase is linear over the bandwidth of the pulse. For opaque barriers these flatness and linearity conditions hold well within the stopband. Tunneling without distortion requires that the pulse bandwidth be narrow compared to the width of the stopband of the barrier. The requirement that $\Omega \ll \Omega_c$ then implies that the pulse width must satisfy the constraint

$$\tau_p \gg \frac{2\pi}{\kappa v}. \quad (27)$$

For a pulse whose amplitude attenuates exponentially as $\exp(-\kappa z)$ within the barrier, the above condition means that

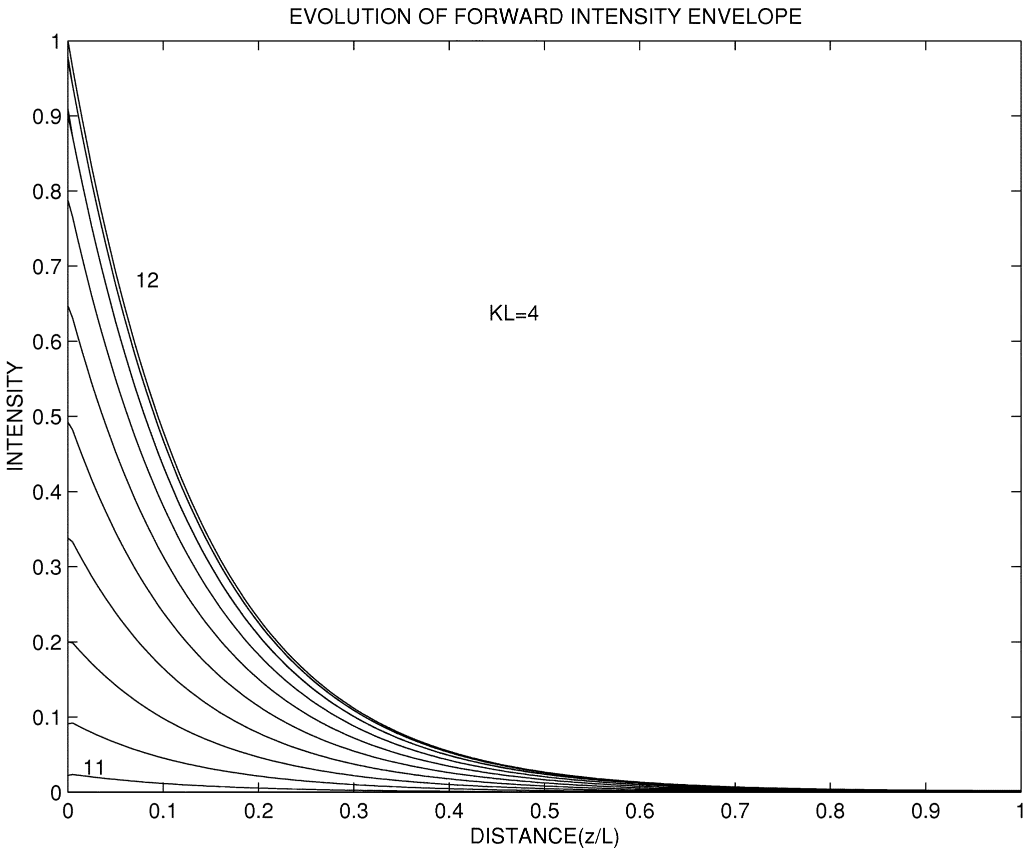

Fig. 6. Evolution of forward intensity envelope over one cycle.

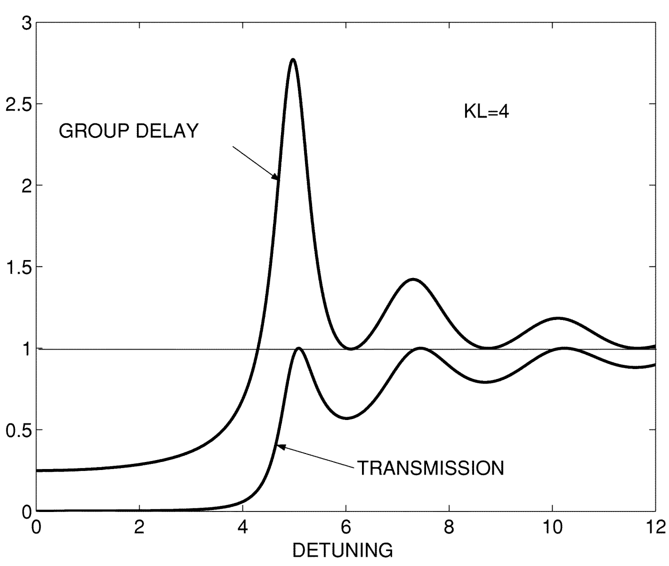

Fig. 7. The transmission and normalized group delay for a PBG of strength $\kappa L = 4$. The normalized detuning parameter is ΩT_{tr}.

the pulse spatial extent $v\tau_p$ must be much more than six times the decay distance $1/\kappa$. In terms of the normalized coupling constant κL and the transit time $T_{tr} = L/v$, the distortionless condition reads $\tau_p \gg 2\pi T_{tr}/\kappa L$. For fairly opaque barriers with reasonable transmission $(1 \leq \kappa L \leq 6)$ the pulse length must exceed the transit time of a light front through the barrier in order to avoid distortion. This is what defines the domain of quasistatics: the pulse duration so long compared to the transit time that over the bulk of the interaction, steady state conditions apply. The barrier acts as a lumped element with respect to the pulse envelope which cannot be localized within the barrier. Experiments that have reported tunneling without pulse distortion have all been done in this regime [5], [8]–[10]. In cases (such as for the thickest samples in [8]) where this condition was violated, significant pulse narrowing was observed.

For narrowband pulses the tunneling process is a quasisteady state phenomenon in which the field envelope throughout the barrier can follow the slow variations of the input envelope with little phase lag. In this quasistatic limit we can obtain approximate solutions to the coupled-mode equations for arbitrary input pulse profiles by expanding the complex amplitudes of the sinusoidal solutions to first order in the frequency parameter $\Omega/\kappa v$ and performing an inverse Fourier transform, whereupon $i\Omega \rightarrow -\partial/\partial t$. The resulting solutions are

$$E_F(z,t) = \frac{\cosh\kappa(z-L)}{\cosh\kappa L} \times \left\{ A(t) - \frac{1}{\kappa v}\left[\tanh\kappa L + \tanh\kappa(z-L)\right] \times A'(t) \right\} \tag{28}$$

$$E_B(z,t) = -i\frac{\sinh\kappa(z-L)}{\cosh\kappa L} \times \left\{ A(t) - \frac{\tanh\kappa L}{\kappa v} A'(t) \right\}. \tag{29}$$

Here, $A(t)$ is the envelope of the incident pulse as measured at $z = 0$ and the primes denote derivatives with respect to time. The first term in the quasistatic expansion is obtained by letting

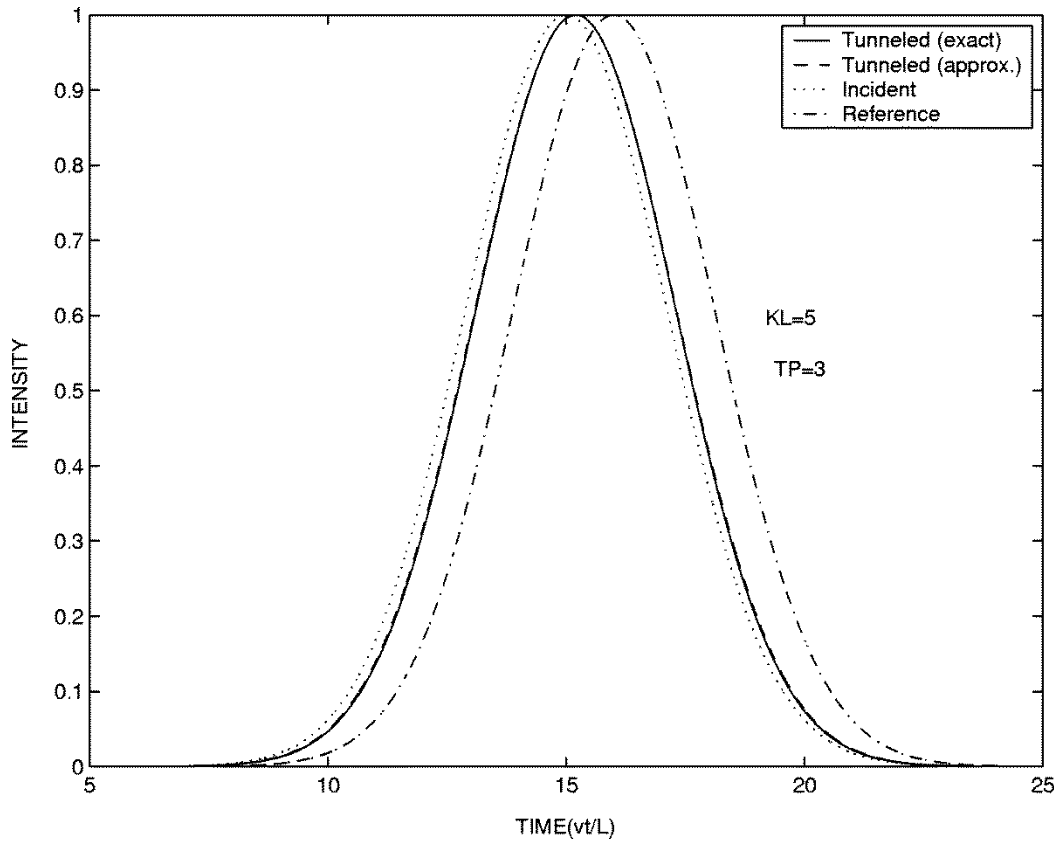

Fig. 8. Incident (dotted), transmitted (solid), and reference (dotted-dashed) pulses for a barrier of strength $\kappa L = 5$. On top of the tunneled pulse and indistinguishable from it is the quasistatic solution.

A be a slow function of time. It is obvious that if the pulse envelope is much longer than the barrier length, over a large part of the pulse the fields in the barrier will simply be the steady-state fields multiplied by the instantaneous envelope amplitude at the input. This leads to the most trivial form of apparent superluminality where the output envelope adiabatically follows the input envelope because both are essentially in steady state. Under these conditions, a peak may be measured simultaneously at the input and output merely because the peak is so broad compared to the barrier length that both the transit time of a light front and the energy storage time are negligible. The next terms in the quasistatic expansion involve time derivatives of the input pulse envelope. The derivative terms are multiplied by factors which are simply the group delays of the corresponding forward and backward waves. The delay for the forward wave actually depends on z as a result of the superposition of cosh and sinh modes discussed in Section VII. However, the only physically measurable delays are the ones at the exit for E_F and at the input for E_B. The delays originate from the coupling-induced energy storage and the impedance mismatch at the boundaries.

We first consider the limiting case of an infinitely long barrier. In that case, the quasistatic fields become

$$E_F(z,t) = e^{-\kappa z} A(t) \tag{30}$$

$$\begin{aligned} E_B(z,t) &= ie^{-\kappa z}\left[A(t) - \left(\frac{1}{\kappa v}\right)A'(t)\right] \\ &\approx ie^{-\kappa z} A\left(t - \frac{1}{\Omega_c}\right). \end{aligned} \tag{31}$$

These results are interesting. They tell us that the forward envelope at every point z follows the incident envelope with no delay. It is an exponentially decaying standing wave, an exponential spatial mode whose amplitude is modulated by the incident pulse envelope. It does not propagate. So long as it does not encounter an exit boundary that extracts energy, it will follow the incident field with no phase lag. The backward wave, on the other hand, is in phase quadrature to the forward wave and is delayed by an amount equal to the inverse of the cutoff angular frequency. This delay time is equal to the time it takes for energy in the pulse to be stored and released by the barrier. Since no energy escapes in the forward direction, the entire delay is observed in the reflected pulse.

A delay in the appearance of a peak is just that, a delay. One must not make the leap from a *delay time* to a velocity. In fact, in an operational sense, one never really measures a velocity. What one measures are delay times and distances. Velocities are always inferred and that requires a knowledge of a trajectory. In lumped circuit engineering it is easy enough to create delay elements and phase shifters that do not involve any propagation effects. Energy storage can result in delays without *propagation*.

As shown in Section IX, these quasistatic solutions are in excellent agreement with the numerical solutions of the coupled mode equations after the initial transient which lasts about two transit times.

IX. Simulations of Gaussian Pulse Tunneling

To follow a tunneling pulse, we integrate the coupled-mode equations along forward and backward characteristics for

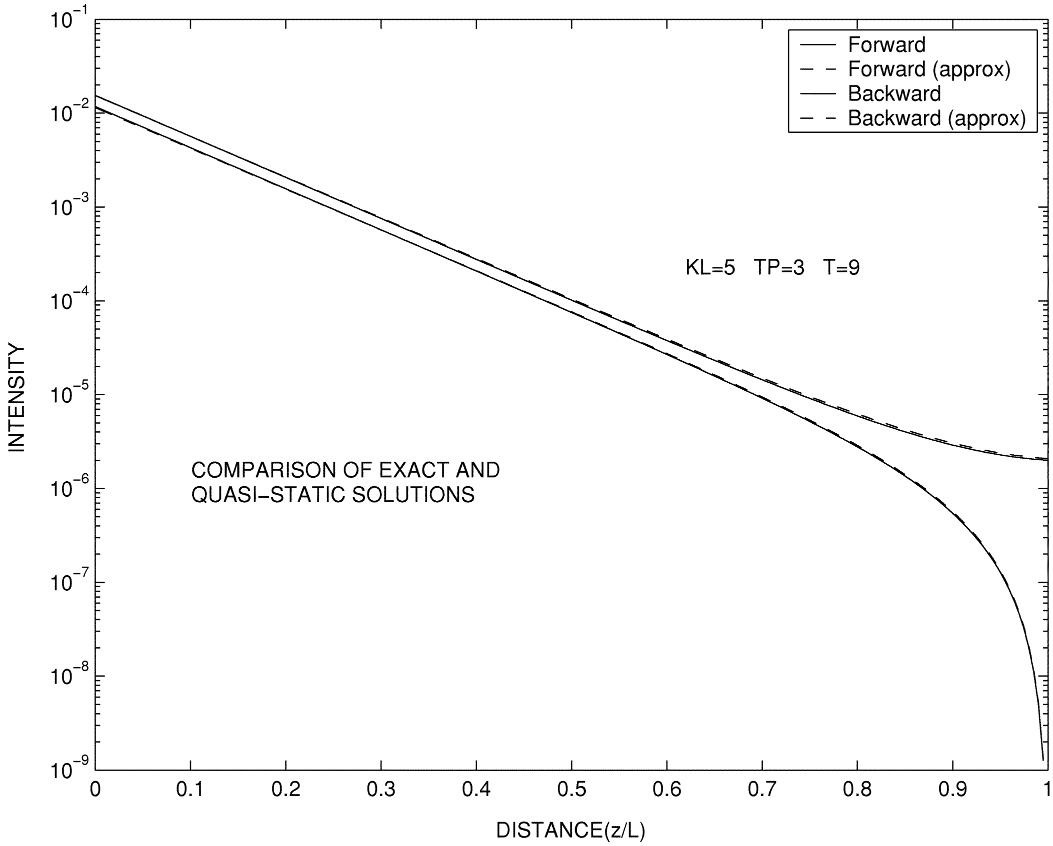

Fig. 9. Comparison of the exact and quasistatic intensity distributions at time $t = 9$.

Gaussian input pulses of the form

$$A(t) = E_f(0,t) = \exp\left[-\frac{(t-t_0)^2}{2\tau_p^2}\right]. \qquad (32)$$

In these simulations time is measured in units of the transit time and t_0 is chosen so that at $t = 0$ the incident field is negligible. Typically, $t_0 = 5\tau_p$. We take $\kappa L = 5$ so that the barrier is quite opaque, with a mid-gap transmission of 1.8×10^{-4}. This is comparable to the transmission of the thickest barriers in the experiment of Spielmann, *et al.* [8]. First, we take a pulse of width $\tau_p = 3$ so that the adiabatic condition is satisfied. Fig. 8 shows the incident pulse, the transmitted (tunneled) pulse, and the reference pulse (i.e., a pulse that travels the distance L in a barrier-free region. The transmitted pulse is normalized by its peak value of 1.8×10^{-4}. The transmitted pulse is undistorted and its peak is delayed by a time $\tau_g \approx 0.2$ with respect to the input peak. This delay, which equals $1/\kappa L$, is exactly the limiting value predicted by the phase time or group delay. The reflected pulse is identical in shape and suffers an equal delay. Overlaid (as a dashed line) on top of the transmitted pulse in Fig. 8 is the quasistatic approximation to the output. It is impossible to distinguish one from the other. The quasistatic field distributions are established within one or two transit times. Fig. 9 compares the exact and quasistatic solutions at the instant $t = 9$, which is still well within the front tail of the input pulse. Over the bulk of the pulse the fields are accurately described by the quasistatic solutions. This confirms our assertion that distortionless tunneling is a quasistatic phenomenon.

The peak of the tunneled pulse appears at the output in one-fifth the time it would take for a pulse traveling at c to traverse an equal distance in the unperturbed medium. However, one must not leap to the conclusion that the pulse tunneled at five times the speed of light. The fact is that the incident peak *did not propagate* to the output. For a narrowband pulse such as this, the peak does not even enter the barrier. Fig. 10 shows snapshots of the intrabarrier energy density $u(z,t) \propto \left[|E_F(z,t)|^2 + |E_B(z,t)|^2\right]$ taken at successive instants between $t = 10$ and $t = 20$, during the rise and fall of the pulse. The pulse energy density is a monotonically decreasing function of z. The entire distribution rises and falls as a semirigid entity in response to the input modulation. The distributions for the descending part of the pulse (dashed curves) lag behind the rising distributions as a result of energy storage. Also, near the exit, the fields deviate from their near-exponential distributions as they adjust to the loading conditions at that end. Clearly, the output peak and input peak are not related by causal propagation. Furthermore, the entire input pulse, and not just its early tails, contributes to the output as it modulates the stored energy. The duration of the tunneling process is just the length of the input pulse. There is no observable reshaping, distortion, or compression of the transmitted pulse.

Tunneling without distortion is only possible if the input pulse is longer than the barrier width. For the thickest sample in the Spielman experiment, this condition was not met and significant pulse shortening was observed. Here, we simulate that situation by taking a pulse of width $\tau_p = 0.6$. Fig. 11 shows that the pulse is shortened and reshaped. The reshaping is due to

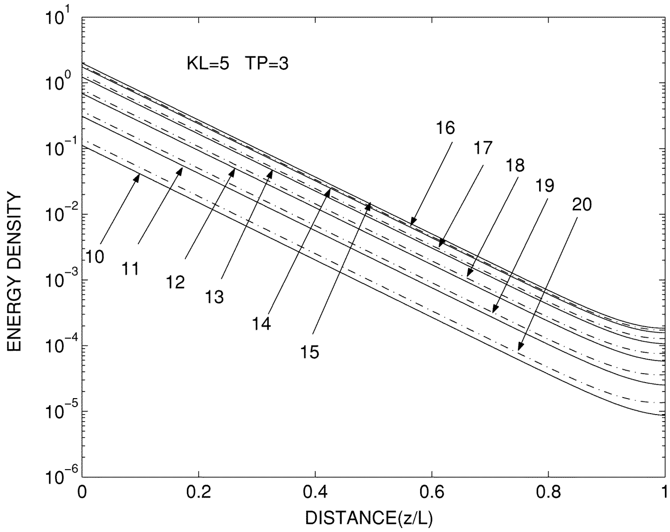

Fig. 10. Spatial distribution of pulse energy density inside barrier.

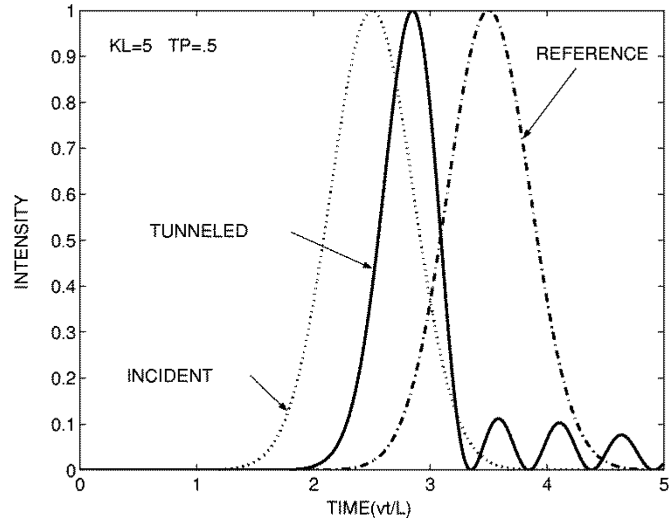

Fig. 12. Pulse breakup seen with a shorter wavepacket. Note ringing at cutoff frequency.

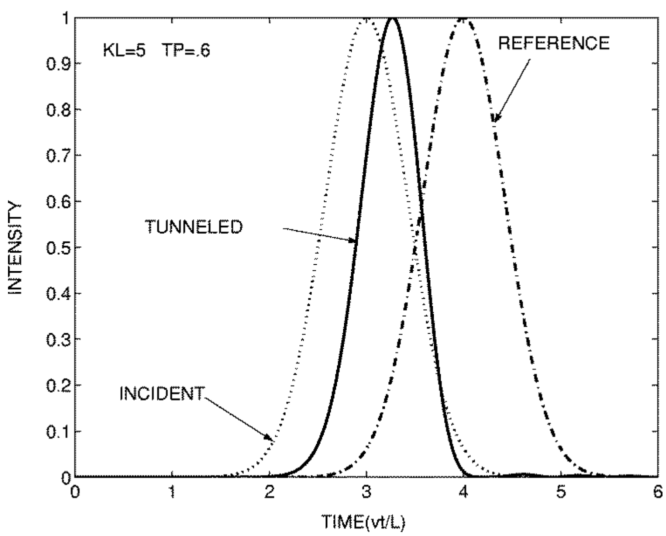

Fig. 11. Tunneling of a shorter pulse. Note pulse shortening and distortion.

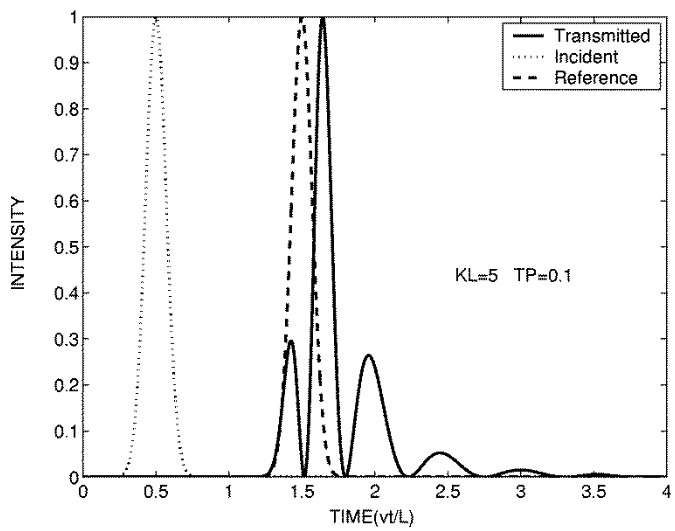

Fig. 13. Transmission of a very short pulse.

group-delay-distortion (GDD) that arises from the higher order terms neglected in the quasistatic expansion. For even shorter pulses (Figs. 12 and 13), we observe pulse breakup and the ringing expected of an impulsively excited cavity. This ringing is at the cutoff frequency. For the shortest input pulse shown here $(\tau_p = 0.1)$ the transmission is luminal. However, this pulse is not actually tunneling but "flying over" the barrier as it has significant spectral content in the filter pass bands. Snapshots of this pulse at different instants of time show a peak in the energy density actually propagating through the barrier (Fig. 14). This is in marked contrast to the behavior of narrowband pulses where the peak does not even enter the barrier.

X. Relation Between Group Delay, Dwell Time, and Stored Energy

The group delay is proportional to the energy stored in the barrier [16]. Within the slowly varying envelope approximation, the group delay for this tunneling problem is also identical to the dwell time, a quantity introduced by Smith to characterize the lifetime of collisions [27]. Here, we demonstrate explicitly that the group delay is directly proportional to the time-averaged stored energy in the barrier and is equal to the dwell time.

The time-averaged stored energy within the barrier is given by

$$\begin{aligned}\langle U\rangle =& \frac{1}{4}\int_{vol}\left[\mu|H|^2+\varepsilon|E|^2\right]dv\\ =& \frac{1}{2}\varepsilon\int_{vol}\left[|E_F|^2+|E_B|^2\right]dv\end{aligned} \tag{33}$$

which yields

$$\langle U\rangle = \left(\frac{1}{2}\varepsilon E_0^2 A\right) \times\left[\frac{\kappa^2}{\gamma^2}\frac{\tanh\gamma L}{\gamma} - L\left(\frac{\Omega}{\gamma v}\right)^2 \operatorname{sech}^2\gamma L\right]\cos^2\phi_t. \tag{34}$$

The time-averaged incident power is

$$P_i = \frac{1}{2}\varepsilon E_0^2 A v. \tag{35}$$

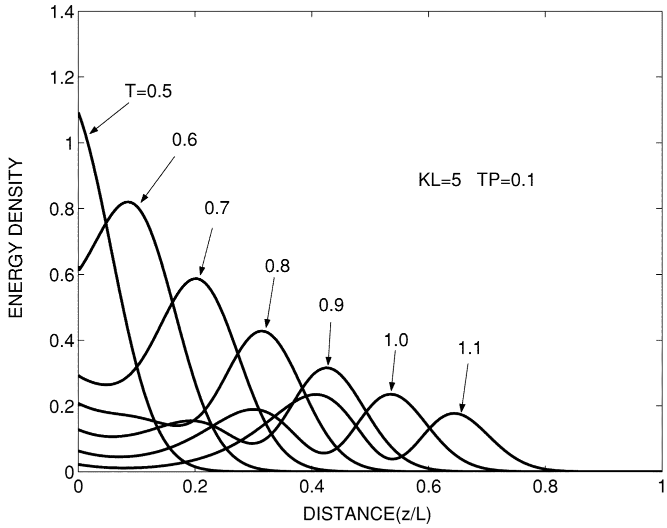

Fig. 14. Spatial profiles of a short "tunneling" pulse. Peak passes through the barrier.

Upon comparing (34) with the expression in (25) for the group delay, we find that the group delay is directly proportional to the time-averaged stored energy

$$\langle U\rangle = P_i\tau_g. \tag{36}$$

On the other hand, the dwell time is defined as

$$\tau_d = \frac{\langle U\rangle}{P_i} \tag{37}$$

and is by definition proportional to the stored energy. Thus, we find that the group delay and the dwell time are identical and are proportional to the stored energy.

This proportionality between stored energy and group delay is a well-known result that has been proven for both lumped-element networks and distributed microwave circuits [28], [29]. Basically, the time required for a pulse of energy to enter an element and leave again is the average stored energy per unit incident cw power. In the context of barrier tunneling it is important not to make the leap from an anomalously short group delay to a superluminal propagation velocity. The delay seen here is due to energy storage and is not a propagation delay. We note that the role of energy storage has been discussed in other superluminal contexts involving absorptive or amplifying media [30]–[32]. There, the energy is stored in the atomic excitations. In the tunneling problem considered here, the energy is stored in the electromagnetic cavity formed by the structure.

XI. Origin of the "Hartman Effect"

As noted by Hartman [11] and confirmed by experiment [8], [12], the group delay becomes independent of length for thick barriers. This would imply arbitrarily large and even infinite velocities for long enough barriers if tunneling was a propagation phenomenon. A physical explanation of this paradoxical result has been lacking. We have recently provided an explanation on the basis of energy storage and the fact that the tunneling wave does not actually propagate but exists throughout the barrier as a standing wave with an exponentially decaying envelope [15], [16]. This envelope adiabatically follows the input pulse envelope with a small time lag given by the group delay. The group delay is proportional to the time averaged stored energy in the barrier. If the stored energy becomes independent of barrier length, then so must the group delay. This is the origin of the "Hartman effect".

From (34) for the stored energy, we find

$$\lim_{L\to\infty}\langle U\rangle = P_i\cdot\left(\frac{1}{\gamma v}\right) \tag{38}$$

which, for $\Omega \ll \kappa v$ is equal to $P_i/\kappa v$. Thus, the stored energy saturates at a value given by the ratio of the input power to the cutoff angular frequency. The group delay then becomes independent of length and is given by the inverse of the cutoff angular frequency. Again, this is not a propagation delay and should not be taken to imply arbitrarily large velocities for increasing barrier length.

The stored energy saturates because the field is a nearly exponentially decreasing function of distance. Beyond a decay distance $1/\kappa$ it does not matter how much more length a barrier has since the energy is practically all stored within this attenuation distance. Of course, for a thick but not quite infinite barrier, a finite but exponentially small amount of energy does reach the exit and is characterized by this saturated delay time. Note, however, that for a truly infinite barrier, the delay refers to the delay of the reflected pulse as discussed in Section X. For the forward field, so long as no exit boundary is encountered, all points are in phase with the input field and do not exhibit any delay.

XII. Discussion

Our conclusion is that tunneling is not a propagation phenomenon but a quasistatic process mediated by a standing wave in the barrier. The anomalously short delays observed in tunneling have their origin in energy storage and its subsequent release. An anomalously short delay does not imply superluminal velocity. While a delay time is always a useful concept, a velocity need not be, especially if the delay is not a propagation delay.

Because it is mediated by a quasistatic field, the tunneling process cannot possibly be used to signal at speeds greater than c. The time scale of changes in the pulse envelope is much longer than the transit time of a light front through the barrier. Information requires change, and nothing changes faster or travels faster than the light front.

Appendix

Global Energetics of Pulse Tunneling

It has been argued that little physical significance attaches to delay times based on peaks [4]. However, these peaks are peaks of energy fluxes and should tell us something about energy flow through and storage within the barrier. In Section IV, we showed how the local energy storage and power flow is determined by the coupling strength κ. We now use the macroscopic Poynting theorem to show that, in fact, the delay in the appearance of reflected and transmitted energy flux peaks is exactly the time it takes for energy to be stored and released by the barrier.

The instantaneous total energy stored in the cavity is

$$U(t) = \int_{vol} u(z,t)dz$$
$$= A\int_0^L \frac{1}{2}\varepsilon\left[|E_F(z,t)|^2 + |E_B(z,t)|^2\right]dz \quad \text{(I.1)}$$

where A is the cross-sectional area of the barrier normal to the z direction. The real instantaneous Poynting vector $\mathbf{E}\times\mathbf{H}$ is given by

$$\mathbf{S}(z,t) = \left(\frac{1}{2\eta}\right)(|E_F(z,t)|^2 - |E_B(z,t)|^2)\hat{\mathbf{z}} \quad \text{(I.2)}$$

and represents the net electromagnetic energy flux. Note that this flux is the difference between forward and backward components, both of which are required for tunneling to occur. Poynting's theorem then tells us that

$$-\oint_s \mathbf{S}\cdot\hat{\mathbf{n}}da = \frac{dU}{dt} \quad \text{(I.3)}$$

where the surface integral on the left hand side reduces to an integration over the input and output faces of the barrier. At the exit surface only the transmitted forward flux ($P_t(t) \propto |E_F(L,t)|^2$) exists. At the input surface the net flux is simply the incident power ($P_i(t) \propto |E_F(0,t)|^2$) minus the reflected power ($P_r(t) \propto |E_B(0,t)|^2$). Thus, evaluation of the surface integral in (I.4) yields

$$P_i(t) - P_r(t) - P_t(t) = \frac{dU}{dt}. \quad \text{(I.4)}$$

This relation simply says that the rate of increase of stored electromagnetic energy in the cavity is given by the incident power minus the sum of the transmitted and reflected power. It is important to note that the balance between incident power on one hand and the sum of reflected and transmitted power on the other hand, $P_i(t) = P_r(t) + P_t)(t)$, does not hold in the transient case as a result of energy storage in the barrier. An intensity peak seen at the output could be the result of the cavity releasing energy it had stored from an earlier time.

For thick barriers the transmitted power near steady state is negligible compared to the incident and reflected powers, hence we can write

$$P_i(t) - P_r(t) = \frac{dU}{dt}. \quad \text{(I.5)}$$

We see from (I.5) that the stored electromagnetic energy in the barrier is increasing when the instantaneous incident power exceeds the reflected power and is decreasing otherwise. It reaches a maximum at the instant when the reflected power equals the incident power. For narrowband pulses we have seen that an incident pulse is reflected without distortion but with a small delay τ_g. For opaque barriers the reflectivity is close to unity and, hence, to an excellent approximation

$$P_r(t) = P_i(t - \tau_g). \quad \text{(I.6)}$$

The time t_M at which the stored barrier energy reaches a maximum can then be found by solving

$$P_i(t_M) = P_i(t_M - \tau_g). \quad \text{(I.7)}$$

For any symmetric pulse with a peak at t_0, one can replace the behavior around the peak by a quadratic in $(t - t_0)$. Then, solving

$$(t_M - t_0)^2 = (t_M - t_0 - \tau_g)^2 \quad \text{(I.8)}$$

we find that the maximum of the total stored energy in the barrier occurs at the instant

$$t_M = t_0 + \frac{\tau_g}{2}. \quad \text{(I.9)}$$

This result tells us that the delay of the reflected pulse peak is twice the delay of the peak of the stored barrier energy. In other words, the group delay is the time it takes for energy to be stored and released by the barrier. The peaks of the reflected and transmitted pulses do indeed have physical significance.

Acknowledgment

The author acknowledges useful conversations with R. Chiao, J. Eberly, G. Nimtz, E. Recami, and other attendees of the Quantum Optics mini-workshop, Kavli Institute for Theoretical Physics, University of California, Santa Barbara.

References

[1] R. Landauer, "Light faster than light?," *Nature*, vol. 365, pp. 692–693, 1993.
[2] R. Y. Chiao and A. M. Steinberg, "Tunneling times and superluminality," in *Progress in Optics*, E. Wolf, Ed. Amsterdam, The Netherlands: Elsevier, 1997, vol. 37, pp. 347–406.
[3] L. A. MacColl, "Note on the transmission and reflection of wave packets by potential barriers," *Phys. Rev.*, vol. 40, pp. 621–626, 1932.
[4] T. Martin and R. Landauer, "Time delay of evanescent electromagnetic waves and analogy to particle tunneling," *Phys. Rev. A*, vol. 45, pp. 2611–2617, 1992.
[5] A. M. Steinberg, P. G. Kwiat, and R. Y. Chiao, "Measurement of the single-photon tunneling time," *Phys. Rev. Lett.*, vol. 71, pp. 708–711, 1993.
[6] Y. Japha and G. Kurizki, "Superluminal delays of coherent pulses in nondissipative media: A universal mechanism," *Phys. Rev. A*, vol. 53, pp. 586–590, 1996.
[7] G. Diener, "Superluminal group velocities and information transfer," *Phys. Lett. A*, vol. 223, pp. 327–331, 1996.
[8] C. Spielmann, R. Szipocs, A. Stingl, and F. Krausz, "Tunneling of optical pulses through photonic bandgaps," *Phys. Rev. Lett.*, vol. 73, pp. 2308–2311, 1994.
[9] M. Mojahedi, E. Schamiloglu, F. Hegeler, and K. J. Malloy, "Time-domain detection of superluminal group velocity for single microwave pulses," *Phys. Rev. E*, vol. 62, pp. 5758–5766, 2000.
[10] S. Longhi, M. Marano, P. Laporta, and M. Belmonte, "Superluminal optical pulse propagation at 1.5 μm in periodic fiber Bragg gratings," *Phys. Rev. E*, vol. 64, pp. 055 602(R) 1–4, 2001.
[11] T. E. Hartman, "Tunneling of a wave packet," *J. Appl. Phys.*, vol. 33, pp. 3427–3433, 1962.
[12] A. Enders and G. Nimtz, "Zero-time tunneling of evanescent mode packets," *J. Phys. I (France)*, vol. 3, p. 1089, 1993.
[13] A. P. L. Barbero, H. E. Hernandez-Figueroa, and E. Recami, *Phys. Rev. E.*, vol. 62, pp. 8628–8635, 2000.
[14] E. H. Hauge and J. A. Stovneng, "Tunneling times: A critical review," *Rev. Mod. Phys.*, vol. 61, pp. 917–936, 1989.
[15] H. G. Winful, "The nature of "superluminal" barrier tunneling," *Phys. Rev. Lett.*, vol. 90, no. 032901, 2003.
[16] ——, "Energy storage and group delay in superluminal barrier tunneling: Origin of the Hartman effect," *Opt. Express*, vol. 10, pp. 1491–1497, 2002.
[17] A. Ranfagni, D. Mugnai, P. Fabeni, and G. P. Pazzi, *Appl. Phys. Lett.*, vol. 58, p. 774, 1991.
[18] A. D. Jackson, A. Lande, and B. Lautrup, "Apparent superluminal behavior in wave propagation," *Phys. Rev. A*, vol. 64, pp. 0 044 101-1–4, 2001.

[19] T. Emig, "Propagation of an electromagnetic pulse through a waveguide with a barrier: A time domain solution within classical electrodynamics," *Phys. Rev. E*, vol. 54, pp. 5780–5787, 1996.
[20] J. M. Deutch and F. E. Low, "Barrier penetration and superluminal velocity," *Ann. Phys. (N.Y.)*, vol. 210, pp. 184–202, 1993.
[21] H. G. Winful, "Pulse compression in optical fiber-filters," *Appl. Phys. Lett.*.
[22] F. S. Crawford, *Waves:Berkeley Physics Course-Volume 3*. New York: McGraw-Hill, 1968.
[23] L. A. Vainshtein, "Propagation of pulses," *Sov. Phys. Usp.*, vol. 19, pp. 189–205, 1976.
[24] A. Carballar and M. A. Miguel, "Phase reconstruction from reflectivity in fiber Bragg gratings," *J. Lightwave Technol.*, vol. 15, pp. 1314–1322, 1997.
[25] B. Costa, D. Mazzoni, M. Puleo, and E. Vezzoni, "Phase shift technique for the measurement of chromatic dispersion in optical fibers using LED's," *IEEE J. Quantum Electron.*, vol. QE-18, pp. 1509–1515, 1982.
[26] S. Ryu, Y. Horiuchi, and K. Mochizuki, "Novel chromatic dispersion measurement method over continuous gigahertz tuning range," *J. Lightwave Technol.*, vol. 7, pp. 1177–1180, 1989.
[27] F. T. Smith, "Lifetime matrix in collision theory," *Phys. Rev.*, vol. 118, pp. 349–356, 1960.
[28] C. G. Montgomery, R. H. Dicke, and E. M. Purcell, *Principles of Microwave Circuits*. New York: McGraw-Hill, 1948.
[29] G. Kishi and K. Nakazawa, "Relations between reactive energy and group delay in lumped-constant networks," *IEEE Trans. Circuit Theory*, pp. 67–71, 1963.
[30] C. G. B. Garrett and D. E. McCumber, "Propagation of a Gaussian light pulse through an anomalous dispersion medium," *Phys. Rev. A*, vol. 1, pp. 305–313, 1970.
[31] G. Diener, "Energy transport in dispersive media and superluminal group velocities," *Phys. Lett. A*, vol. 235, pp. 118–124, 1997.
[32] M. Ware, S. A. Glasgow, and J. Peatross, "Role of group velocity in tracking field energy in linear dielectrics," *Opt. Express*, vol. 9, p. 519, 2001.

Herbert G. Winful (M'76–SM'81–F'87) was born in London, U.K., on December 3, 1952. He received the SB degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, in 1975 and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, in 1981.

From 1980 to 1986, he was a Member of Technical Staff at GTE Laboratories, Waltham, MA, where he conducted research in nonlinear optics, semiconductor laser physics, and fiber optics. He joined the Electrical Engineering and Computer Science Department at the University of Michigan, Ann Arbor, in 1987. His current research interests include nonlinear phenomena in photonic bandgap structures and the physics of tunneling through potential barriers.

Dr. Winful is a Fellow of the American Physical Society and the Optical Society of America. He has won numerous teaching awards, including the State of Michigan Teaching Excellence Award, the Arthur F. Thurnau Professorship, the Amoco/University Teaching Excellence Award, the College of Engineering Teaching Excellence Award, the EECS Professor of the Year Award (twice), the College of Engineering Outstanding Professor Award, and the EECS Teaching Excellence Award. He was named a Presidential Young Investigator in 1986. For eight years he served as faculty advisor to Eta Kappa Nu, the electrical engineering honor society. He is a member of Eta Kappa Nu and Tau Beta Pi.