Tunneling of a Wave Packet

Cite as: Journal of Applied Physics **33**, 3427 (1962); https://doi.org/10.1063/1.1702424
Submitted: 17 May 1962 . Published Online: 11 June 2004

Thomas E. Hartman

Tunneling of a Wave Packet

THOMAS E. HARTMAN
Texas Instruments Incorporated, Dallas 22, Texas
(Received May 17, 1962)

Analytic expressions are given for the time spent by a particle tunneling through a potential barrier. The expressions are derived for an incident wave packet which is initially Gaussian, centered about a point an arbitrary distance away from a rectangular potential barrier and moving toward the barrier with constant average velocity. Upon collision with the barrier, the packet splits into a transmitted and a reflected packet. The resultant transmission time is positive, nonzero and in principle measurable. Although the transmission time becomes quite large as the incident kinetic energy becomes very small, in general, for nonzero incident momentum and finite potential barriers which are neither very thick nor very thin, the transmission times are less than the time that would be required for the incident particle to travel a distance equal to the barrier thickness. The transmission times for metal-insulator-metal thin film sandwiches, given approximately by

$$\delta t_3 = \hbar/(E_f\phi)^{\frac{1}{2}},$$

where E_f is the Fermi energy of the metal, and ϕ the vacuum work function, are of the order of 10^{-16} sec, compared to RC time constants of about 10^{-13} sec.

1. INTRODUCTION

RECENTLY, considerable interest has developed in tunneling across thin insulating layers[1–4] and the possible application of this phenomenon to practical devices.[5–9] Stationary energy state tunneling through potential barriers has been treated extensively[10]; however, the time dependent problem has not received much attention.

To account for the time spent by a tunneling particle in the classically forbidden region, the transmission of a wave packet must be considered. The requirement for such a solution was originally suggested by Condon[11] in 1931 and discussed the following year by MacColl.[12] An elegant solution, limited to the case of a potential barrier having the form of a δ function has been reported by Jánossy.[13]

The proposed frequency limitations of devices involving the tunneling process have been based on the dielectric relaxation time[14] and resistance-capacitance time constants,[6] where it has been stated that tunneling takes place in an "extremely short" time. While discussing tunnel diodes, Sommers[15] has pointed out the continued need for a wave packet analysis and states that the time required classically for an electron moving with thermal velocity to travel a distance equal to the barrier thickness is of the same order of magnitude as the dielectric relaxation time but as he suggests, it is doubtful if this has any real significance.

Discussion of the time dependent behavior of a wave packet is complicated by its diffuse or spreading nature; however, the position of the peak of a symmetrical packet can be described with some precision. When the incident packet collides with the barrier it splits into a transmitted and a reflected packet. MacColl states that the peak of the transmitted packet leaves the barrier at "about the instant" the peak of the incident packet arrives at the barrier, so that there is no "appreciable" delay. MacColl has restricted his consideration to a packet which is composed of plane wave components corresponding to energies well away from the top (and bottom) of the barrier. In this way any subsequent increase in the probability of finding the particle in the transmitted region can be attributed solely to tunneling.

In this paper a general solution is presented for the simplest possible model which can be solved exactly, i.e., the passage of a one-dimensional Gaussian wave packet through a rectangular potential barrier. The incident packet is composed of plane wave components corresponding to all kinetic energies greater than zero. For very thick barriers most of the tunneling takes place very near the top of the barrier. Therefore, general results can be obtained only for a packet whose energy distribution has not been terminated an arbitrary amount below the top of the barrier.

The solutions are constructed by the time-honored technique of finding the stationary state solutions, appending the appropriate time factor together with the initial position condition and Gaussian weighting function and then integrating over all the possible states.[16]

The more significant results can be summarized by

[1] J. C. Fisher and I. Giaever, J. Appl. Phys. **32**, 172 (1961).
[2] C. A. Mead, Phys. Rev. Letters **6**, 545 (1961); **8**, 56 (1962).
[3] I. Giaever, Phys. Rev. Letters **5**, 147, 464 (1960).
[4] J. Nicol, S. Shapiro and P. H. Smith, Phys. Rev. Letters **5**, 461 (1960).
[5] C. A. Mead, Proc. IRE **48**, 359, 1478 (1960).
[6] C. A. Mead, J. Appl. Phys. **32**, 646 (1961).
[7] J. P. Spratt, R. F. Schwarz, and W. M. Kane, Phys. Rev. Letters **6**, 341 (1961).
[8] E. Burstein, D. N. Langenberg, and B. N. Taylor, Phys. Rev. Letters **6**, 92 (1961).
[9] R. N. Hall, Solid-State Electronics **3**, 320 (1961).
[10] For example, see: L. deBroglie, Ann. Inst. Henri Poincaré **3**, 349 (1933); and M. Datzeff, Ann. Phys. (Paris) **10**, 583 (1938).
[11] E. U. Condon, Revs. Modern Phys. **3**, 43 (1931).
[12] L. A. MacColl, Phys. Rev. **40**, 621 (1932).
[13] L. Jánossy, Acta. Phys. Acad. Hung. Sci. **2**, 171 (1952).
[14] A. G. Chynoweth, *Progress in Semiconductors*, edited by A. F. Gibson (John Wiley & Sons, Inc., New York, 1960), Vol. 4, p. 97.
[15] H. S. Sommers, Jr., Proc. IRE **47**, 1201 (1959).
[16] D. Bohm, *Quantum Theory* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1951), Chaps. 3 and 11.

comparing the transmission time to the time required for the incident packet to traverse a distance equal to the barrier thickness. The latter time will be referred to as the "equal time." For very thin barriers the transmitted packet has substantially the same form as the incident packet in agreement with the results of Jánossy. The resultant transmission time is longer than the "equal time." For thicker barriers the peak of the transmitted packet is shifted, relative to the incident packet, to higher energy values. The transmission time becomes independent of barrier thickness and small compared to the "equal time." For very thick barriers the form of the transmitted packet is badly distorted with the greatest contribution coming from Fourier components corresponding to energies just above the top of the barrier where the transit time is approximately the "equal time."

2. ANALYSIS

The rectangular potential barrier is represented by the potential energy function defined for the three different regions by the following equations

$$\begin{aligned} V(x)&=0, & x&\leqslant 0, & &\text{(region I)};\\ &=V_0, & 0&<x<a, & &\text{(region II)};\\ &=0, & a&\leqslant x, & &\text{(region III)}. \end{aligned} \tag{1}$$

Region I is the incident (and reflected) region, region II is the barrier and region III is the transmitted region.

The one-dimensional time-dependent Schrödinger equation is

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

and the following wave packet solutions are obtained:

$$\begin{aligned} \Psi(x,t)&=\varphi_1(x,t)+\chi_1(x,t), & x&\leqslant 0;\\ &=\varphi_2(x,t)+\chi_2(x,t), & 0&<x<a;\\ &=\varphi_3(x,t), & a&\leqslant x; \end{aligned} \tag{3}$$

where

$$\varphi_1(x,t)=\frac{1}{\Delta k(2\pi)^{\frac{1}{2}}}\int_0^\infty f(k_1-k_0)\exp i\left[k_1(x+x_0)-\frac{E(k_1)t}{\hbar}\right]dk_1,$$

$$\chi_1(x,t)=\frac{1}{\Delta k(2\pi)^{\frac{1}{2}}}\int_0^\epsilon f(k_1-k_0)B_1(k_1)\exp i\left[k_1(-x+x_0)+\beta_1(k_1)-\frac{E(k_1)t}{\hbar}\right]dk_1$$
$$+\frac{1}{\Delta k(2\pi)^{\frac{1}{2}}}\int_\epsilon^\infty f(k_1-k_0)D_1(k_1)\exp i\left[k_1(-x+x_0)+\delta_1(k_1)-\frac{E(k_1)t}{\hbar}\right]dk_1,$$

$$\varphi_2(x,t)=\frac{1}{\Delta k(2\pi)^{\frac{1}{2}}}\int_0^\epsilon f(k_1-k_0)A_2(k_1)\exp\left[k_2'x+ik_1x_0+i\alpha_2(k_1)-\frac{iE(k_1)t}{\hbar}\right]dk_1$$
$$+\frac{1}{\Delta k(2\pi)^{\frac{1}{2}}}\int_\epsilon^\infty f(k_1-k_0)C_2(k_1)\exp i\left[k_2x+k_1x_0+\gamma_2(k_1)-\frac{E(k_1)t}{\hbar}\right]dk_1, \tag{4}$$

$$\chi_2(x,t)=\frac{1}{\Delta k(2\pi)^{\frac{1}{2}}}\int_0^\epsilon f(k_1-k_0)B_2(k_1)\exp\left[-k_2'x+ik_1x_0+i\beta_2(k_1)-\frac{iE(k_1)t}{\hbar}\right]dk_1$$
$$+\frac{1}{\Delta k(2\pi)^{\frac{1}{2}}}\int_\epsilon^\infty f(k_1-k_0)D_2(k_1)\exp i\left[-k_2x+k_1x_0+\delta_2(k_1)-\frac{E(k_1)t}{\hbar}\right]dk_1,$$

$$\varphi_3(x,t)=\frac{1}{\Delta k(2\pi)^{\frac{1}{2}}}\int_0^\epsilon f(k_1-k_0)A_3(k_1)\exp i\left[k_1(x+x_0)+\alpha_3(k_1)-\frac{E(k_1)t}{\hbar}\right]dk_1$$
$$+\frac{1}{\Delta k(2\pi)^{\frac{1}{2}}}\int_\epsilon^\infty f(k_1-k_0)C_3(k_1)\exp i\left[k_1(x+x_0)+\gamma_3(k_1)-\frac{E(k_1)t}{\hbar}\right]dk_1;$$

where

$$\begin{aligned} A_3(k_1)&=2k_1k_2'\cdot F(k_1),\\ B_1(k_1)&=(k_1^2+k'^2_2)\sinh(ak_2')\cdot F(k_1),\\ C_3(k_1)&=2k_1k_2\cdot G(k_1),\\ D_1(k_1)&=(k_1^2-k_2^2)\sin(ak_2)\cdot G(k_1),\\ F(k_1)&=[4k_1^2k'^2_2\cosh^2(ak_2')\\ &\quad+(k_1^2-k'^2_2)^2\sinh^2(ak_2')]^{-\frac{1}{2}}, \end{aligned} \tag{5}$$

$$\begin{aligned} G(k_1)&=[4k_1^2k_2^2\cos^2(ak_2)+(k_1^2+k_2^2)\sin^2(ak_2)]^{-\frac{1}{2}},\\ \alpha_3(k_1)&=\arctan\{[(k_1^2-k'^2_2)/2k_1k_2']\tanh(ak_2')\}-ak_1,\\ \beta_1(k_1)&=\arctan\{[-2k_1k_2'/(k_1^2-k'^2_2)]\coth(ak_2')\},\\ \gamma_3(k_1)&=\arctan\{[(k_1^2+k_2^2)/2k_1k_2]\tan(ak_2)\}-ak_1,\\ \delta_1(k_1)&=\arctan\{[-2k_1k_2/(k_1^2+k_2^2)]\cot(ak_2)\}; \end{aligned}$$

and

$$
\begin{aligned}
k_1 &= [2mE(k_1)/\hbar^2]^{\frac{1}{2}},\\
k_2 &= \{2m[E(k_1)-V_0]/\hbar^2\}^{\frac{1}{2}} = [k_1^2-\epsilon^2]^{\frac{1}{2}},\\
k_2' &= \{2m[V_0-E(k_1)]/\hbar^2\}^{\frac{1}{2}} = [\epsilon^2-k_1^2]^{\frac{1}{2}}, \qquad (6)\\
\epsilon &= [2mV_0/\hbar^2]^{\frac{1}{2}};
\end{aligned}
$$

and $E(k_1)$ is the total energy, $\hbar$ is Planck's constant divided by 2π, m is the mass of the particle, and ϵ is the height of the barrier in wavenumbers. The coefficients are complex functions of k_1 and have been written in the polar form where A, B, C, and D are the moduli and α, β, γ, and δ are the phases. The coefficients of the incident wave have arbitrarily been normalized. The Gaussian weighting function is given as

$$f(k_1-k_0)=\exp[-(k_1-k_0)^2/2(\Delta k)^2], \qquad (7)$$

where Δk characterizes the spread in energy values associated with the packet peaked at an average total energy E_0 corresponding to

$$k_0=[2mE_0/\hbar^2]^{\frac{1}{2}}$$

It is easy to verify that the function $\Psi(x,t)$ [Eq. (3)] satisfies Eq. (2) where $V(x)$ is given by Eq. (1) and that for any fixed value of t, Ψ and $\partial\Psi/\partial x$ are continuous functions of x.

The function φ_1 is the incident wave packet; χ_1 is the reflected wave packet; and φ_3 is the transmitted wave packet. The integrations in Eqs. (4) are difficult to perform and the integrals have not been evaluated, but considerable insight into the physical behavior of the system can be obtained by investigating the nature of the modulus and phase of the integrand for the transmitted packet. The magnitude of the contribution of each plane wave component is determined by the modulus. This contribution will be appreciable only in the neighborhood of space and time where the phase is changing slowly with respect to k_1.[16]

2.1 Modulus

The modulus of the transmitted packet integrand which pertains to the plane wave components with $k_1 \leqslant \epsilon$ is $f(k_1-k_0)A_3(k_1)$. The function $A_3(k_1)$ is the square root of the transmission coefficient and is plotted as a function of incident wavenumber in Fig. 1, where the parameter is barrier thickness a in units of $1/\epsilon$. The graph of the Gaussian weighting function $f(k_1-k_0)$ [Eq. (7)] for $k_0=0.7\epsilon$ and $\Delta k=0.1k_0$ is also shown.

For very thin barriers (i.e., $a\epsilon \ll 1$), $A_3(k_1)$ is nearly constant and almost unity (except for $k_1 \ll \epsilon$). Hence for k_0/ϵ not too near zero, the function $f(k_1-k_0)A_3(k_1)$ has a maximum near k_0 and has substantially the same form as the Gaussian incident packet.[17]

For larger values of a, however, the transmission coefficient remains small except for k_0/ϵ near unity where its square root increases more rapidly than the weighting function $f(k_1-k_0)$ decreases. The product of

[17] Jánossy[13] has found the latter result to hold for the case of a δ-function potential barrier also.

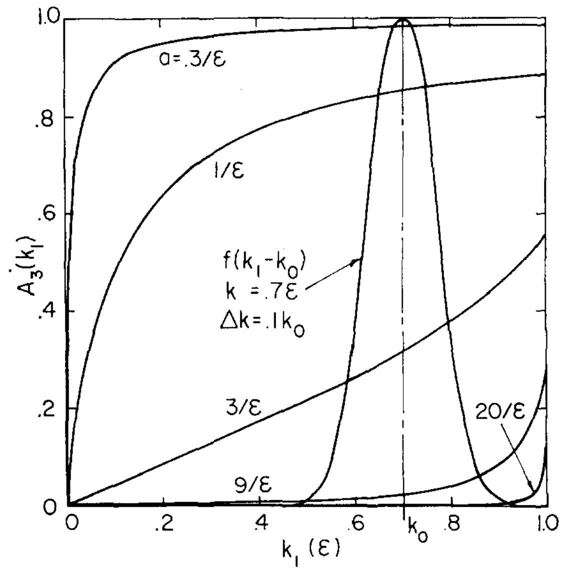

FIG. 1. A graph of the square root of the transmission coefficient $A_3(k_1)$ and the Gaussian weighting function $f(k_1-k_0)$ as functions of incident wavenumber k_1. The parameter a is the barrier thickness.

the two functions results in a maximum at a k_1 somewhat larger than k_0 and a minimum near the k_1 for which $A_3(k_1)$ begins to increase more rapidly than $f(k_1-k_0)$ decreases. For still larger values of a, the transmission coefficient remains small but is appreciably different from zero only for k_1/ϵ near unity. Thus, the function $f(k_1-k_0)A_3(k_1)$ is appreciable only for k_1/ϵ near unity and increases monotonically for all $k_1<\epsilon$. Hence, no maximum (or minimum) exists in the region $k_1<\epsilon$.

For smaller values of Δk than the one shown in Fig. 1, the contribution of higher energy plane wave components decreases more rapidly as k_1 deviates from k_0. Therefore, much larger values of a are required to cause $A_3(k_1)$ to increase rapidly enough to result in a minimum in the modulus.

The relative contribution of each incident plane wave component to the transmitted packet is shown in Figs. 2. For all cases shown $\Delta k=0.1k_0$. Figures 2(a) and (b) are for $k_0=0.3\epsilon$ and 0.5ϵ, respectively. The peak of the transmitted packet shifts to only slightly higher values of k_1 as a increases and the transmitted packet remains essentially Gaussian. Figure 2(c) is for $k_0=0.7\epsilon$ where the range of a values is large enough to display the distortions discussed above. Figure 2(d) is for $k_0=0.9\epsilon$ where the distortions occur for even smaller values of a.

From Figs. 2(c) and (d) it is seen that for sufficiently thick barriers the majority of the tunneling takes place just below the top of the barrier. Thus for many cases it is necessary to consider that portion of the transmitted packet which is composed of plane wave components corresponding to energies greater than the barrier.

The modulus of the transmitted packet integrand which pertains to the plane wave components with

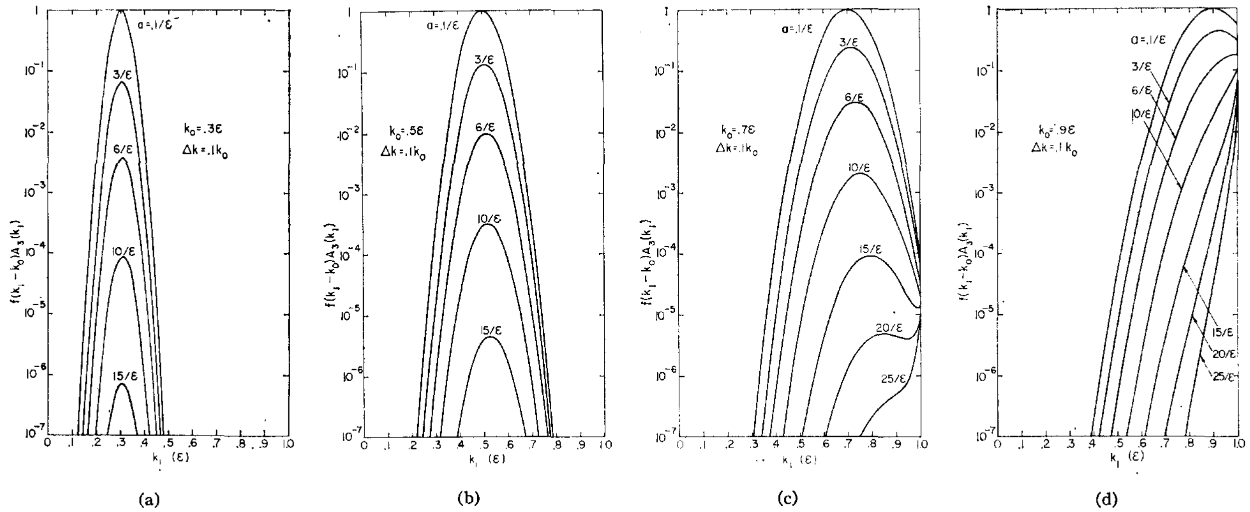

FIG. 2. Graphs of the modulus of the transmitted packet integrands for plane wave components corresponding to energies less than the top of the barrier as a function of incident wavenumber. The average incident wavenumber for (a) is $k_0=0.3\epsilon$, (b) is $k_0=0.5\epsilon$, (c) is $k_0=0.7\epsilon$, and (d) is $k_0=0.9\epsilon$. In all cases $\Delta k=0.1k_0$. The parameter a is the barrier thickness.

$k_1\geqslant\epsilon$ is $f(k_1-k_0)C_3(k_1)$. Again the function $C_3(k_1)$ is the square root of the transmission coefficient, but above the barrier the transmission coefficient is oscillatory. It is easy to show from Eqs. (9) and (11) that $A_3(k_1)$ and $C_3(k_1)$ form a continuous function in k_1 and that for $k_1\gg\epsilon$, $C_3(k_1)\approx 1$. The function $C_3(k_1)$ oscillates between a steadily increasing lower envelope and unity. A graph of $C_3(k_1)$ for $a=4/\epsilon$ is shown in Fig. 3. There is perfect transmission when $ak_2=n\pi$ where $n=1, 2, 3, \cdots$, i.e., whenever the barrier thickness is equal to an integral number of half wavelengths of the plane wave component. The values of k_1 for which $C_3(k_1)=1$ are given by

$$k_1=(\epsilon^2+n^2\pi^2/a^2)^{\frac{1}{2}};\quad n=1, 2, 3, \cdots.$$

The oscillations in $C_3(k_1)$ have been used to account for observed periodic deviations from the Schottky emission plot.[18]

The relative contribution of each incident plane wave component to the transmitted packet is $f(k_1-k_0)A_3(k_1)$ for $k_1\leqslant\epsilon$ and $f(k_1-k_0)C_3(k_1)$ for $k_1\geqslant\epsilon$. A graph of the relative contribution is shown in Fig. 4 for the case of $k_0=0.9\epsilon$, $\Delta k=0.1k_0$, and $a=10/\epsilon$. The curve a is a graph of $f(k_1-k_0)$ for $k_0=0.9\epsilon$ and is also the upper bound for the graph of $f(k_1-k_0)C_3(k_1)$. At the k_1 values which result in a transmission coefficient of unity the graph of the function $f(k_1-k_0)C_3(k_1)$ will become tangent to the curve a. A lower bound for $f(k_1-k_0)C_3(k_1)$ in the region $k_1\geqslant\epsilon$ is formed by the mirror image of $f(k_1-k_0)A_3(k_1)$. The greatest contribution to the transmitted packet is seen to come from slightly above the barrier. For very thick barriers the contribution just above the barrier is further accentuated and the vast majority of the tunneling occurs just below the barrier.[19]

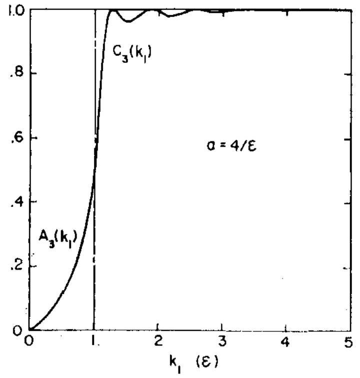

FIG. 3. A graph of the square root of the transmission coefficient as a function of incident wave-number. (a) for $k_1<\epsilon$ the graph is of $A_3(k_1)$; (b) for $k_1\geqslant\epsilon$ the graph is for $C_3(k_1)$.

[18] E. S. Guth and C. J. Mullin, Phys. Rev. 59, 575 (1941).

2.2 Phase

At any time t a particular plane wave component will contribute appreciably to the packet only in the neighborhood of the x for which the phase of the integrand for that particular component is changing slowly[16] (i.e., at an extremum). From Eq. (4) the condition for an extremum in the phase of the incident packet is

$$\partial\{k_1(x-x_0)-[E(k_1)t/\hbar]\}/\partial k_1=0, \tag{8}$$

and of the transmitted packet when $k_1\leqslant\epsilon$ is

$$\partial\{k_1(x+x_0)+\alpha_3(k_1)-[E(k_1)t/\hbar]\}/\partial k_1=0. \tag{9}$$

Equations (8) and (9) together with Eqs. (5) and (6) result in the following equations of motion for the Fourier component corresponding to k_1 in the incident wave packet

$$x=-x_0+(\hbar k_1/m)t, \tag{10}$$

and the transmitted wave packet[20]

$$x=-x_0-(\partial\alpha_3/\partial k_1)+(\hbar k_1/m)t. \tag{11}$$

[19] Here tunneling is defined as penetration of the barrier proper.

[20] The present analysis differs from that of MacColl where only the ak_1 portion of the argument $\alpha_3(k_1)$ is taken into account. The

The probability that the incident particle has the momentum $\hbar k_1$ can be found using the incident wave packet where the expectation value is $\hbar k_0$. According to the correspondence principle, when $k_1=k_0$, Eq. (10) must describe classically the motion of a particle with momentum $\hbar k_0$ moving in a region of constant potential which is at $x=-x_0$ when $t=0$. Clearly Eq. (10) does this for each plane wave component of the incident packet. The time when a particle with incident momentum $\hbar k_1$ would arrive at the front of the barrier (i.e., $x=0$) is just $mx_0/\hbar k_1$. The time when the same particle, if it is transmitted, would leave the rear of the barrier (i.e., $x=a$) from Eq. (11) is

$$t=\frac{mx_0}{\hbar k_1}+\frac{m}{\hbar k_1}\left[\left(\frac{\partial\alpha_3}{\partial k_1}\right)+a\right],\quad k_1\leqslant\epsilon. \tag{12}$$

The transmission time δt_3 is the time required for a particle with incident momentum $\hbar k_1$ to tunnel and from Eq. (12) is seen to be

$$\delta t_3=(m/\hbar k_1)[(\partial\alpha_3/\partial k_1)+a],\quad k_1\leqslant\epsilon, \tag{13}$$

where

$$\frac{\partial\alpha_3}{\partial k_1}=\frac{1}{k_2'}\left[\frac{\epsilon^4\sinh(2ak_2')-2ak_1^2k_2'(k_1^2-k_2'^2)}{4k_1^2k_2'^2\cosh^2(ak_2')+(k_1^2-k_2'^2)^2\sinh^2(ak_2')}\right]-a.$$

It is just the fact that the contributions from different plane wave components of a wave packet travel with different average velocities that causes the wave packet to spread. Since the plane wave components corresponding to larger momenta will arrive at $x=0$ ahead of the peak of the incident packet (which travels with velocity $\hbar k_0/m$), then it is not surprising to find, in the cases where the peak value of the transmitted packet has been shifted to somewhat higher values of k_1, that the peak of the transmitted packet leaves $x=a$ before the peak of the incident packet arrives at $x=0$. Such a result is not a violation of the principle of causality. It must be remembered that the wave packet can be interpreted as describing the probability that an incident particle has the momentum $\hbar k_1$. When it has the momentum $\hbar k_1$, where $k_1>k_0$, it will arrive at $x=0$ ahead of one which has the more probable momentum $\hbar k_0$ and also left $x=-x_0$ at $t=0$. In any event the particle with momentum $\hbar k_1$ will require on the average the time δt_3 to penetrate the barrier.

The δt_3 for $k_1\geqslant\epsilon$ is found in the same way as δt_3 in Eq. (13) and results in

$$\delta t_3=\frac{m}{\hbar k_1}\left[\left(\frac{\partial\gamma_3}{\partial k_1}\right)+a\right],\quad k_1\geqslant\epsilon, \tag{14}$$

where

$$\frac{\partial\gamma_3}{\partial k_1}=\frac{1}{k_2}\left[\frac{2ak_1^2k_2(k_1^2+k_2^2)-\epsilon^4\sin(2ak_2)}{4k_1^2k_2^2\cos^2(ak_2)+(k_1^2+k_2^2)^2\sin^2(ak_2)}\right]-a.$$

Graphs of the transmission time vs barrier thickness as functions of incident wave number are shown in Fig. 5 as the δt_3 surface. The $\delta t=ma/\hbar k_1$ lines are plots of the "equal time" as functions of incident wave-

remainder was neglected on the grounds that it is slowly varying over the range of interest. All of $\alpha_3(k_1)$ is included in the present analysis since the total magnitude of the time delay is of interest as well as the change in the delay.

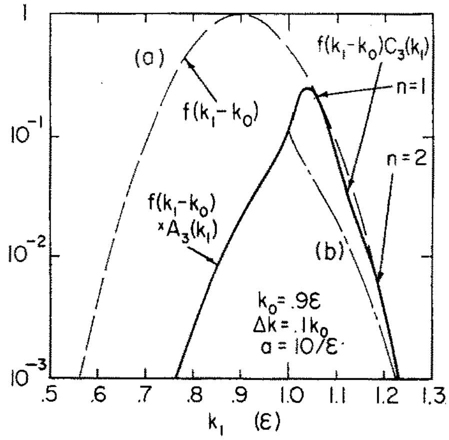

FIG. 4. A graph of the modulus of the transmitted packet integrand for $k_0=0.9\epsilon$, $\Delta k=0.1\,k_0$, and $a=10/\epsilon$. The upper bound (a) is $f(k_1-k_0)$. The lower bound in the region $k_1\geqslant\epsilon$ is the mirror image of $f(k_1-k_0)A_3(k_1)$ about the axis $k_1=\epsilon$. The transmission is perfect for $k_1=(\epsilon^2+n^2\pi^2/a^2)^{\frac{1}{2}}$.

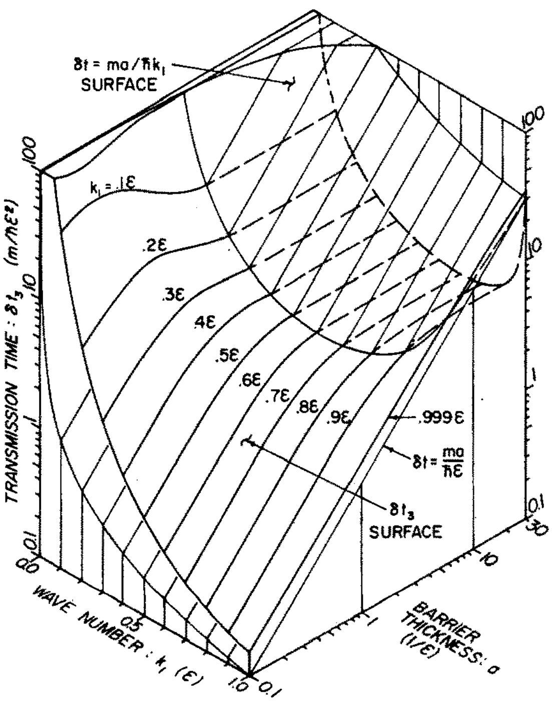

FIG. 5. Graphs of the transmission time δt_3 as a function of barrier thickness a are shown as the δt_3 surface. The δt surface is composed of graphs of the "equal time." For both sets of graphs the parameter k_1 is the incident wavenumber.

number. The part of the transmission time surface which is to the left of the intersection with the "equal time" surface corresponds to a transmission time which is greater than the "equal time." This is referred to as the thin barrier region. The part of the surface which is to the right of the intersection corresponds to the thicker barrier region where the transmission time approaches $2m/\hbar k_1k_2'$ as a limit. For very thick barriers most of the tunneling corresponds to $k_1\approx\epsilon$ because of packet distortion. The equal time for $k_1=\epsilon$ is plotted as $\delta t=ma/\hbar\epsilon$. The transmission time curve for $k_1=0.999\epsilon$ is seen to be approximately parallel to the "equal time" lines over the range of a plotted. The δt_3 curve converges to the $ma/\hbar\epsilon$ line as k_1 approaches ϵ.

The function δt_3 for $k_1\geqslant\epsilon$ given in Eq. (14) is oscillatory and although not plotted in Fig. (5) it is easy to show that the "equal time" surface forms a lower bound for δt_3 and the $ma/\hbar(k_1-\epsilon)$ surface forms an upper bound and for values of $k_1\geqslant\epsilon$ the function δt_3 oscillates between these two bounds. Also for $k_1\gg\epsilon$, $\delta t_3\approx ma/\hbar k_1$ as expected.

It should be noted that for any fixed a and for $\epsilon\gg k_1$ then $\epsilon\approx k_2'$ and from Eq. (13)

$$\delta t_3\approx 2m/\hbar k_1k_2', \tag{13'}$$

therefore the limit as k_1 approaches zero is

$$\lim_{k_1\to 0}\delta t_3=\infty,$$

i.e., as the incident kinetic energy approaches zero the time delay becomes infinite. On the other hand, for a and k_1 fixed

$$\lim_{V_0\to\infty}\delta t_3=0.$$

But from Eq. (5) the magnitude of the transmitted packet (i.e., the modulus) vanishes in the limit of $V_0=\infty$. Therefore perfect reflection does not result from an infinite time delay but rather from a vanishing transmission coefficient.

In addition the present analysis is applicable to the case $V_0=0$. The transmission in this case is described entirely by the latter integral in each of Eqs. (4). The time delay where $k_1=k_2$ and $\epsilon=0$ (i.e., $V_0=0$) is given by Eq. (14) as $\delta t_3=ma/\hbar k_1$, i.e., the transmission time is the "equal time" as would be expected.

Finally, for a diminishingly thin barrier the transmission time is given by Eq. (13) when $ak_2\ll\epsilon$ as

$$\delta t_3=\frac{2m\epsilon^4ak_2'}{\hbar k_1k_2'[4k_1^2k_2'^2+a^2k_2'^2(k_1^2-k_2'^2)^2]};$$

therefore for k_1 and ϵ fixed

$$\lim_{a\to 0}\delta t_3=0.$$

There are only two cases where δt_3 is identically zero. In the former case, $V_0=\infty$, and the magnitude of the transmitted packet vanishes; in the latter case, $a=0$, i.e., there is no barrier. Tunneling is not present in either case, and it is therefore concluded that tunneling is accompanied by a time delay.

3. UNCERTAINTY PRINCIPLE

There are two complementary limits for the canonical variables E and t that are compatible with the uncertainty principle. For this purpose, the uncertainty principle can be written directly in two convenient forms

$$\Delta E\Delta t\geqslant\hbar/2 \tag{15}$$

and

$$\Delta x\Delta k_1\geqslant\tfrac{1}{2}.$$

Upon substitution of Eq. (6), Eq. (15) can be rewritten as

$$\Delta t\geqslant m/2\hbar k_1\Delta k_1. \tag{16}$$

The first complementary limit is $\Delta E\ll E$ or $\Delta k_1\ll k_1$. Let Δk_1 be small enough that

$$\Delta x\geqslant 1/2\Delta k_1\gg a,$$

then from Eq. (16)

$$\Delta t\gg ma/\hbar k_1,$$

i.e., for $\Delta k_1/k_1$ sufficiently small

$$\delta t_3\leqslant\Delta t,$$

as illustrated in Fig. 6. In this limit the incident packet is sharply peaked at $k_1=k_0$. In addition, for a not too large, the transmitted packet is sharply peaked at $k_1=k_0+\delta k$, where $\delta k\ll k_0$. Under these conditions the motion of the transmitted particle can be satisfactorily approximated by the behavior of the Fourier component corresponding to $k_1=k_0+\delta k$. Therefore the transmission time $\delta t_3(k_1)$, where $k_1=k_0+\delta k$ is also characteristic of the particle. The probability that the transmitted particle will arrive at x', where $x'>a$, at the time t is shown schematically in Fig. 6 as P_t. The dotted curve is the probability when $a=0$ (i.e., no barrier present). Therefore this probability is centered about

$$t'=m(x'+x_0)/[\hbar(k_0+\delta k)],$$

and from Eq. (13)

$$(\delta t_3)'=\delta t_3(k_1)=(m/\hbar k_1)[(\partial\alpha_3/\partial k_1)+a],\quad k_1=k_0+\delta k.$$

The most probable time of arrival is given approximately as

$$\bar{t}=t'+(\delta t_3)'\pm[\Delta t+\tau(x')]/2,$$

where $\tau(x')$ is a function which characterizes the spread of the packet in going from $-x_0$ to x'. It should be noted that even though $(\delta t_3)'$ may be small compared to Δt, if the time is measured for a sufficiently large number of identical systems, t can be determined to any arbitrary precision, and because of the correspondence principle $(\delta t_3)'$ can be deduced.

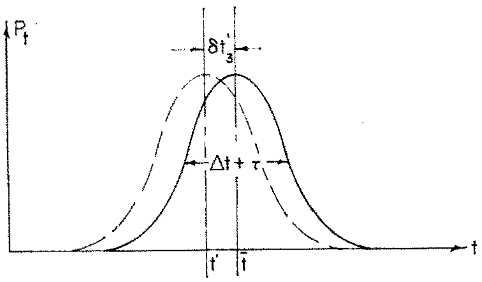

FIG. 6. The probability that a transmitted particle will arrive at x' at the time t is shown schematically as the solid curve. The dashed curve is the probability when $a=0$.

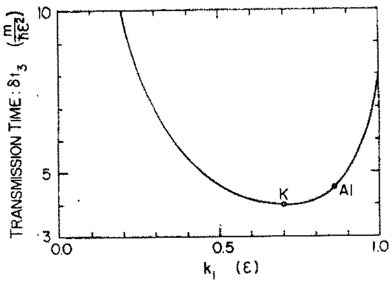

FIG. 7. A graph of the transmission time δt_3 for a 20-Å thick vacuum barrier as a function of incident wavenumber. The two extremes from the k_1/ϵ values calculated for several metal-vacuum-metal sandwiches are plotted on the curve.

The other complementary limit is $\Delta t \ll t$. In this limit Δk_1 is so large that appreciable distortion in the transmitted packet is inevitable. The transmission time no longer can be characterized by a single δt_3 as in the previous limit but must be characterized by a distribution of δt_3's corresponding to all the Fourier components and weighted according to their respective contributions. However, for barriers which are very thick, the majority of the transmission can be characterized by $\delta t_3 \approx ma/\hbar\epsilon$.

The vast majority of cases which are of practical interest lie between these two extremes. In most cases the transmitted packet is substantially symmetrical and the transmission time can be characterized by $\delta t_3(k_1)$, where $k_1=k_0+\delta k$.

4. CONCLUSIONS

It has been shown that the transmission time is positive, nonzero (except for $a=0$ or $V_0=\infty$) and in principle measurable even in the $\Delta k_1 \ll k_0$ limit. For $a<1/\epsilon$, $\delta t_3 > ma/\hbar k_0$, i.e., the transmission time for thin barriers is greater than the time required for the incident packet to traverse a distance equal to the barrier thickness.

For thicker barriers the form of the transmitted packet is substantially the same as the incident packet but centered about $k_1=k_0+\delta k$. The transmission time is independent of barrier thickness and may be given from Eq. (13) where $ak_2 \gg 1$ as

$$\delta t_3 \approx 2m/\hbar(k_0+\delta k)[\epsilon^2-(k_0+\delta k)^2]^{\frac{1}{2}}.$$

The transmission for very thick barriers is over the top of the barrier for the most part, essentially not tunneling at all. The transmission time in this case is given approximately by $\delta t_3 = ma/\hbar\epsilon$.

4.1 Metal-Insulator-Metal Sandwiches

Using the calculated free electron Fermi energies[21] and experimentally measured values for the work functions,[22] δt_3 was calculated for sandwiches with vacuum as an insulating layer of 20-Å thickness. Neglecting image forces and using the same metal on each side of the sandwich, the transmission times for Al, Sb, Be, Bi, Cd, Ce, Cr, Co, Cu, Ga, Au, Pb, Li, Mg, Mn, Mo, Ni, Pt, K, Rh, Ag, Na, Ta, Te, Tl, Sn, Ti, W, V, Zn, and Zr range from 0.95 to 3.01×10^{-16} sec. For all these examples δt_3 is independent of a (for a greater than 3 to 5 Å) and can be given from Eq. (13′) for $V_0=E_f+\phi$ and $E_0=E_f$ as

$$\delta t_3 \approx \hbar/(E_f\phi)^{\frac{1}{2}},$$

where ϕ is the work function of the metal and E_f is the Fermi energy. It is interesting to note that the barrier thickness practical for tunneling devices, Fermi energies, and work functions, for all the examples given above, result in transmission times which lie between that for potassium and aluminum and very near the minimum in the δt_3 surface as shown in Fig. 7.

This fortuitous coincidence all but eliminates the transmission time delay from the list of possible fundamental limitations to frequency response of metal-insulator-metal sandwich tunneling devices compared to the RC time constant. The minimum RC time constant for tunnel emission sandwiches is about 10^{-12} sec.[6] In the case of the tunnel diode, frequency response is probably limited by dielectric relaxation. Highly doped germanium, for example, has a dielectric relaxation time of about 10^{-13} sec.[15]

ACKNOWLEDGMENTS

The author acknowledges with thanks the many helpful and stimulating discussions with M. de Wit and A. R. Beattie, the advice and suggestions of R. Stratton, N. G. Einspruch, and C. A. Barlow, the numerical work done by D. R. Powell, and the critical reading of the manuscript by Chun C. Lin.

[21] C. Kittel, *Introduction to Solid State Physics* (John Wiley & Sons, Inc., New York, 1956), 2nd ed., Chap. 10.

[22] Preferred values from *Handbook of Chemistry and Physics* (Chemical Rubber Publishing Company, Cleveland, Ohio, 1954), 36th ed., p. 2342.