From Photon to Neuron Princeton University Press, 2017

Supplementary Chapter Philip Nelson August 15, 2018

This chapter belongs to Part III (Advanced Topics). It is made freely available as-is in the hope that it will be useful. All reports of errata will be gratefully received: pcn@upenn.edu.

Copyright ©2018 Philip C. Nelson

chapter 16

Tunneling of Photons and Electrons

Don't expect that mathematics will give you an easy answer to any physical problem. If you find that it does, that is evidence that you picked a poor problem to begin with. — Blair Kinsman

16.1 SIGNPOST: INCOMPLETE CANCELLATION

Chapter 5 described the phenomenon of total internal reflection: When light arrives at an interface between media with angle greater than the critical angle, then it cannot penetrate infinitely deeply into the less-dense medium. But then Section 5.3.4 stated that TIR is not "total" after all: Photons can travel a short distance beyond the interface. That is, the "oscillation to death" that supposedly forbids them to escape may be *incomplete*.¹ This chapter will explore this phenomenon, then turn to an analogous one involving electrons. Each situation involves particles appearing where naively they "ought not" to be, so we will call them both **tunneling**. The analysis uses math ideas from complex variables theory, and physics ideas from Chapter 12.

16.2 TUNNELING OF PHOTONS

16.2.1 Setup

Figure 16.1 shows a situation for which ray optics predicts that no light will escape from one medium into another. However, Section 5.3.4 (page 194) stated that actually, photons will be observed at the point indicated, but with probability that falls rapidly with increasing depth h (the "evanescent wave"). We wish to understand that claim.

A point light source is embedded deep in a transparent medium ("water") with index $n_{\rm w}$ at the location shown, where the angle θ exceeds the critical angle² $\theta_{\rm c} = \sin^{-1}(1/n_{\rm w})$. Defining $b = n_{\rm w} \sin \theta$, we therefore are interested in situations with b > 1.

The law of refraction tells us³ that light can emerge into the upper medium along any path whose angle of incidence is less than the critical angle. Such rays intersect the interface at various horizontal distances |u| from the observer, but the figure shows that those distances must all be greater than

 $W = d + L_0(\sin\theta - \tan\theta_c\cos\theta) = d + L_0\cos\theta(\tan\theta - \tan\theta_c).$

Consider the case of a source deep inside the medium, that is, the limit $L_0 \to \infty$ holding θ and d fixed. Because $\theta > \theta_c$, the distance $W \to \infty$, so all stationary-phase

¹Section 4.7.1 (page 161) introduced this notion.

 $^{^{2}}$ See Equation 5.11 (page 194).

³Section 5.3.3 (page 192) introduced the law of refraction.



Figure 16.1: [Schematic.] Total internal reflection setup, a simplification of Figure 5.11 (page 195). An interface is shown between a medium (such as water) with index $n_w > 1$ (*lower*) and air with $n \approx 1$ (*upper*). A point light source (*lower left*) deep inside the medium would normally send light through the interface via stationary-phase paths (*dashed lines*), but in this case such paths are blocked by an opaque barrier. The last stationary phase path exits at horizontal distance W from the observation point (*upper right*), but the barrier extends past this point. Thus, the paths that arrive at an observation point (*solid line*) are all *not* stationary-phase. The path shown corresponds to a negative value of u.

paths are blocked by the opaque barrier. Nevertheless, we may expect that some light would escape into the upper region, because the stationary-phase approximation is not exact. This phenomenon is sometimes called "frustrated total internal reflection" or **photon tunneling**.⁴

Our experience with diffraction suggests that light will not penetrate very far into the "forbidden" region, but Section 5.3.4 pointed out that this can be a virtue: Creating a narrow "evanescent wave" zone where fluorophores can be excited is the basis of TIRF microscopy's high noise rejection.

To investigate, as usual we will consider only paths that consist of two straight-line segments: from source to interface, and from interface to observer. Also as usual, we will drop the slowly-varying modulus factors 1/L in the Light Hypothesis; our concern is the oscillatory behavior that we expect will make the probability amplitude very small. (Of course, every emitted photon must arrive *somewhere*. We are neglecting the possibility of reflection from the interface; later, we will argue that reflection dominates over transmission in the situation shown in the figure.)

Thus, we are interested in the two path segment lengths

$$L_2 = \sqrt{u^2 + h^2}$$

and

$$L_1 = \sqrt{(L_0 \sin \theta + d + u)^2 + {L_0}^2 \cos^2 \theta}.$$

Contents Index Notation

⁴Remarkably, the first observation goes back to Newton, who noticed in the Newton rings phenomenon that the apparent area of contact between the flat and curved glass plates was larger than the actual geometric contact region.

The probability amplitude is a constant times

$$\int_{-d}^{\infty} \mathrm{d}u \, \exp\left[\frac{2\pi \mathrm{i}}{\lambda/n_{\mathrm{w}}}L_1 + \frac{2\pi \mathrm{i}}{\lambda}L_2\right]. \tag{16.1}$$

16.2.2 A complex stationary point controls tunneling probability

In principle, we could evaluate the integral any way we like, for example, numerically. Instead, let's develop an approximate form that yields some insights analogous to the ones we found earlier for diffraction.⁵

In earlier chapters, we often could simplify path integrals by making a stationaryphase approximation. Although we were warned not to expect any stationary-phase points in the present situation, let's look a bit more closely.

In the limit of large L_0 we can simplify $L_1 \approx L_0 + (d+u)\sin\theta$. The phase of the integrand in Equation 16.1 then becomes

$$\varphi \approx \frac{2\pi}{\lambda} \left(u n_{\rm w} \sin \theta + \sqrt{u^2 + h^2} \right) + K,$$
 (16.2)

where K is a real constant (independent of u). Recall from Section 16.2.1 that we define $b = n_{\rm w} \sin \theta$ and that b > 1. For u_* to describe a stationary phase path, we would need

$$0 = b + \frac{u_*}{\sqrt{{u_*}^2 + h^2}},\tag{16.3}$$

that is,

$$u_*{}^2 = h^2/(b^{-2} - 1). (16.4)$$

Because we are assuming that the right hand side is negative, there is no real value u_* that satisfies this condition; as expected, the integral is oscillatory.

The key observation is now that the integrand in Equation 16.1 can be regarded as a complex function evaluated on the real u axis, and that function is analytic except for branch cuts on the imaginary axis at $|u| \ge h$. We can therefore evaluate it by integrating on any convenient contour obtained by deforming the one we have been using. If there are stationary-phase points *anywhere* on the complex u plane, we can use them to approximate our integral.⁶ Equation 16.4 implies that indeed there are such points on the imaginary axes.

We need to be precise about the meaning of square root functions in our formulas. For real values of u there is no problem: We always want the positive square root. This condition implies that the branch cuts starting at $u = \pm ih$ must extend outward, away from the real axis (Figure 16.2). To be explicit, we define $F(u) = G(u^2 + h^2)$ where G is the usual complex square root, defined by

$$G(re^{i\alpha}) = \sqrt{r}e^{i\alpha/2}$$
 where $-\pi < \alpha < \pi$.

For example, $G(-1 \pm i\epsilon) \rightarrow \pm i$ as $\epsilon \rightarrow 0$ through real positive values. Thus, at the point labeled p_+ on the figure, $F(p_+)$ has positive imaginary part, and the opposite for p_- . From now on, we will only use the radical symbol for the square root of a real, positive quantity, where it has its usual meaning.

 $^{^5 {\}rm The}$ analysis in this section follows MV Berry, unpublished (2017), who is gratefully acknowledged. $^6 {\rm This}$ method is discussed in Mathews & Walker, 1970, §3–6.



Figure 16.2: [Sketch.] Contour deformation in the complex u plane. The cross indicates the stationary point u_{*-} (Equation 16.5). Hatched parts of the imaginary axis indicate branch cuts of the function $\varphi(u)$ in Equation 16.2.

With this specification, we can see that the derivative of F is F'(u) = u/F(u), and the condition for stationarity is that $b + u_*/F(u_*) = 0$. Let's look for solutions in the upper half u plane. Equation 16.4 suggests that we check the points

$$u_{*\pm} = \frac{\mathrm{i}h}{\sqrt{1 - b^{-2}}} \pm \epsilon.$$
(16.5)

The last term indicates that we are not sure yet which side of the cut we will choose; the notation " $+\epsilon$ " indicates a limit approaching the imaginary axis from the right. Substituting into the stationarity condition gives

$$\frac{\mathrm{i}h}{\sqrt{1-b^{-2}}} = -b(\pm\mathrm{i})\sqrt{-\left(\frac{-h^2}{1-b^{-2}}+h^2\right)} = \mp\mathrm{i}bh\sqrt{\frac{b^{-2}}{1-b^{-2}}}$$

We see that only the lower sign works, so we have a single stationary point u_{*-} in the upper half u plane.

We now ask if we can deform the integration contour in a way that passes through the stationary-phase point. Examining the phase function φ (Equation 16.2) shows that if u has negative imaginary part, then $e^{i\varphi}$ will be a growing exponential, and hence potentially large. Such a contour would give a result with significant contributions all along its length, which would not be a useful simplification. However, if we deform the contour into the upper u half-plane, then this problem does not arise. Figure 16.2 shows such a contour chosen to pass through the stationary point. We may then expect that the entire integral is dominated by the contribution from the neighborhood of that point, analogously to the case in ray optics.⁷ The key difference from ray optics is that we cannot ignore the constant part of the phase, because it is *not real*. Instead, we get an overall multiplicative factor of $e^{i\varphi}$, where

$$\begin{split} \varphi_* &= \frac{2\pi}{\lambda} \left(\frac{\mathrm{i}bh}{\sqrt{1 - b^{-2}}} - \mathrm{i}\sqrt{-\left(\frac{-h^2}{1 - b^{-2}} + h^2\right)} \right) \\ &= \mathrm{i}\frac{2\pi bh}{\lambda} \sqrt{1 - b^{-2}}. \end{split}$$

Thus $e^{i\varphi_*}$ is indeed a real factor, not a complex phase with modulus 1 as in the derivations in Chapter 6. Although it appears as an overall factor for the transmission

⁷This method is often called "saddle-point approximation"; again see Mathews & Walker, 1970, §3–6.



Figure 16.3: Fluorescence induced by photon tunneling. (a) [Schematics.] Fabrication of a thin film. A layer one molecule thick of the fatty acid $C_{20}H_{40}0_2$ was spread at an air-water interface. A glass slide was repeatedly dipped into the water. When the slide is withdrawn (*left*), the hydrophilic parts of each molecule in the surface layer adhere to the wet glass and one monomolecular layer is deposited. By another immersion of the now hydrophobic slide (*center*) a second fatty-acid layer adheres, and by withdrawal (*right*) a third layer is deposited. The procedure can be repeated again and again to build up a multilayer system consisting of hundreds of monolayers. Two special layers containing fluorophore molecules (*red asterisk*) were deposited, followed by 16 additional, nonfluorescent layers, resulting in the fluorophores' being buried at a fixed distance from the surface. (b) [Experimental data and fit.] The film fabricated as in (a) was illuminated by 476 nm light, and its fluorescence intensity was recorded at various angles of incidence, in order to probe the probability of photons penetrating 45 nm below the surface. The *solid curve* is the prediction from photon tunneling theory. [(a): After Drexhage, 1970. (b): Data from Carniglia et al., 1972.]

probability amplitude, it is not present in the corresponding reflection amplitude. Thus it suppresses transmission relative to reflection.⁸ Its modulus squared takes the form e^{-h/h_*} , where $h_* = (4\pi/\lambda)\sqrt{(n_w \sin \theta)^2 - 1}$.

We have found that the probability for a photon to be observed at distance h from the interface *falls exponentially with* h, with an e-folding length scale given by h_* . The first few e-folding lengths form a layer that was called the "evanescent field" region in Section 5.3.4.

16.2.3 An experiment on fluorescence in the evanescent-field region

C. Carniglia and coauthors developed an experiment that created a planar layer containing fluorescent molecules and was located at a precise distance z = 45 nm from the surface of a transparent medium (Figure 16.3a). They immersed this layered structure in a fluid with index of refraction greater by a factor of 1.10. Light shone through the fluid was then totally internally reflected when its angle of incidence exceeded $\theta_{\text{crit}} = \sin^{-1}(1/1.10)$.

Although z was fixed, the experimenters varied the angle of incidence while observing the fluorescence intensity. To predict its form, they used a calculation method similar to that in Chapter 13. This approach yielded a more detailed formula than the one given in this chapter: In addition to the exponential factor, the authors found

⁸We have neglected the usual transmission and reflection factors (Idea 5.1). These factors are order unity and do not depend on the distance h.



Figure 16.4: [Path diagram.] Electron tunneling. An electron emitted from a source at position x = -A (asterisk) encounters a barrier (blue), at which it may either bounce (dashed line) or pass through to a detector at x = +L (top right).

an additional $\cos^2 \theta$, which made the falloff even steeper. Figure 16.3b shows this prediction alongside the experimental data.

16.3 TUNNELING OF ELECTRONS

16.3.1 Setup

The preceding discussion introduced a mathematical framework that is also useful to analyze a physically quite different situation. Suppose that an electron encounters a barrier. A simple model for such a barrier could be a plane of positive charge, closely followed by a neutralizing plane of negative charge. Outside the double charge layer, the electron feels zero net force; between the layers, it feels a force in the -x direction. Thus, the thin double layer sets up a jump in potential energy by some positive amount U for an electron trying to cross to positive x.

In classical mechanics, an electron arriving with kinetic energy less than U must bounce back from the barrier. Our experience with total internal reflection, however, suggests that something more interesting could happen:⁹ Classical mechanics is only a stationary-phase approximation, so there may be a nonzero probability to detect an electron in the classically forbidden region x > 0. In analogy to Section 16.2.1, we will call this phenomenon **electron tunneling**.

To get started, we must implement the idea of a "monochromatic" (fixed-energy) source of electrons. We expect that, for energy to be fixed, the *time* of emission must be uncertain. Just as an excited atom or molecule emits photons at an indeterminate time, so will the heated filament of the electron source in an electron microscope, an electron donor molecule in a transport chain, and so on. By analogy with Equation 12.24 (page 392), we therefore propose

A monochromatic electron source with energy E is a subsystem whose probability amplitude Ψ_{emis} for emitting an electron at time t_{e} equals a constant times $\exp(-iEt_{\text{e}}/\hbar)$. That is, Ψ_{emis} oscillates with frequency $E/(2\pi\hbar)$. (16.6)

⁹See Section 5.3.4 (page 194).

16.3.2 A complex stationary point again controls tunneling probability

We now use ideas from Sections 12.2.2–12.2.3 to find an approximate formula for the probability amplitude to detect electrons with energy E < U at a point beyond the barrier. Figure 16.4 shows a barrier at fixed position x = 0, detector at fixed x = L, and source at fixed x = -A. The detector is turned on briefly at time t = 0, but the emission time $t = t_e$ is variable. The time at which the electron encounters the barrier, t = -B, is also variable, and not directly observed.

We can classify each path contributing to this process by the values of t_e and B. Each path consists of two segments. The sum over such paths factorizes into the product of two propagator functions.¹⁰ As in Section 16.2.1, we now replace each full propagator by just its integrand (a phase) evaluated on its stationary-phase path. The phase is given as usual by the action functional:

$$S[\text{path}] = \frac{m_{\text{e}}}{2} \int dt \left\| \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} \right\|^2.$$
 [12.1, page 383]

Combining the source phase in Idea 16.6 with Equation 12.1 then gives an expression for the probability amplitude:

$$\int_{-\infty}^{0} \mathrm{d}t_{\mathrm{e}} \,\mathrm{e}^{-\mathrm{i}Et_{\mathrm{e}}/\hbar} \int_{\mathrm{paths}} \mathrm{e}^{\mathrm{i}S[\mathrm{path}]/\hbar}$$
$$= \int_{-\infty}^{0} \mathrm{d}t_{\mathrm{e}} \,\mathrm{e}^{-\mathrm{i}Et_{\mathrm{e}}/\hbar} \int \mathrm{d}B \,\exp\left[\frac{\mathrm{i}m_{\mathrm{e}}}{2\hbar} \frac{A^{2}}{-t_{\mathrm{e}}-B}\right] \exp\left[\frac{\mathrm{i}}{2\hbar} \left(\frac{m_{\mathrm{e}}L^{2}}{B} - 2UB\right)\right]. \tag{16.7}$$

We now check whether there is any set of t_e and B values that constitute a stationaryphase point of the integrand. That would require that two conditions be met:

$$0 = E - \frac{m_{\rm e}A^2}{2(-t_{\rm e}-B)^2} \quad \text{and} \quad 0 = \frac{m_{\rm e}A^2}{2(-t_{\rm e}-B)^2} - \frac{m_{\rm e}L^2}{2B^2} - U.$$
(16.8)

If we were holding t_e fixed, then we would only have the second of these equations, which always has a solution regardless how large we take the barrier U to be. With fixed *energy*, however, Equations 16.8 give solutions

$$B_{*\pm} = \pm \sqrt{m_{\rm e} L^2 / (2(E-U))} \tag{16.9}$$

that are not real when E < U. In that case there is no stationary-phase path, the integral is oscillatory, and the probability amplitude for transmission is small, as expected.

To quantify *how* small, we now recall from Section 16.2.2 that a stationary point may be used to estimate an integral, even if it is off the real axis. Substituting Equation 16.9 into the integrand of Equation 16.7 gives the exponential of

$$\frac{\mathrm{i}}{\hbar} \left[E \sqrt{\frac{m_{\mathrm{e}} A^2}{2E}} + EB_{*\pm} + \frac{m_{\mathrm{e}} A^2}{2} \sqrt{\frac{2E}{m_{\mathrm{e}} A^2}} + \frac{m_{\mathrm{e}}}{2} \frac{L^2}{B_{*\pm}} - UB_{*\pm} \right].$$

We are interested in the real part of this expression, because the imaginary part, when exponentiated, gives a factor with modulus one. The real part is $\pm L\sqrt{2m_{\rm e}(E-U)}/\hbar$.

 $^{^{10}}$ The approach here is similar to that taken in Section 12.2.4 (page 384). The propagator was defined in Equation 12.4.



Figure 16.5: [Experimental data.] Distance dependence of tunneling conductance. Both the probe and the surface were made of gold. The absolute tip-surface distance was not known, so an arbitrary shift has been applied to the horizontal axis. [Data courtesy Ruslan Temirov, Forschungszentrum Jülich.]

As in photon tunneling, the negative solution controls the probability amplitude, which we find to be nonzero but exponentially suppressed as we move more deeply into the forbidden region. Taking the modulus squared gives a probability for tunneling that is

$$\exp\left[-2L\sqrt{2m_{\rm e}(E-U)}/\hbar\right],\tag{16.10}$$

times a slowly-varying prefactor.

Equation 16.10 predicts that the probability for tunneling varies rapidly with distance L. This property led to the invention of **scanning tunneling microscopy**, in which a finely pointed probe is scanned close to, but not quite touching, the surface of an object of interest. Figure 16.5 shows that the probability of electron transport, and hence the current across a junction at fixed potential difference, indeed varies exponentially with gap distance.

16.3.3 Electron tunneling in biomolecules

Prior to the 1970s, chemists believed that reactants that were stable in isolation generally had to collide before a reaction could occur. Stability meant for example that the isolated molecules held their electrons indefinitely; each electron was bound in a potential energy well too deep for it to surmount. When two such molecules were close but not physically touching, they remained stable for the same reason. An exception was possible if the molecules were held in a flexible arrangement: Then thermal fluctuations could momentarily bring them into contact, allowing a reaction such as electron transfer to proceed.

The preceding reasoning led to the expectation that electron transport, such as that between molecules of the respiratory and photosynthetic transport chains, would proceed at a rate proportional to the fraction of time spent in the required conformation. According to Section 0.6, that fraction (a probability) should approach zero as temperature approaches absolute zero. More precisely, Equation 1.20 (page 53) predicts the exponential dependence $\exp(-E_{\text{barrier}}/k_{\text{B}}T)$, that is,¹¹

The logarithm of the rate of a thermally-activated reaction should be linearly proportional to 1/T. (16.11)

 $^{^{11}{\}rm More}$ detailed calculations acknowledge a distribution of many activation barriers as the surrounding molecules flex in thermal motion



Figure 16.6: [Experimental data.] DeVault and Chance's historic evidence for electron tunneling in biomolecules. Temperature dependence of the rate of an electron transfer reaction in *Chromatium*, a purple sulfur bacterium capable of photosynthesis. Distance along the horizontal axis is proportional to T^{-1} . Distance along the vertical axis is proportional to the logarithm of reaction rate. [Data from DeVault et al., 1967.]

A plot of log(rate) versus 1/T is called an **Arrhenius plot**. Figure 16.6 shows that Idea 16.11 is realized at ordinary temperatures for an important biomolecular electron transfer reaction. However, D. DeVault and B. Chance found that below 100 K, instead of falling to zero, the transfer rate leveled off: Some process other than thermal motion was also contributing to electron transfer.

The picture that emerged for this phenomenon involves *two* distinct kinds of tunneling phenomena. First, the electron in question must itself traverse an energetic barrier, because the donor and acceptor never literally touch. The rate for this tunneling is small, but it can be enhanced if the energies of the electronic states before and after the jump match.¹² Those energies, in turn, depend on the conformations of the molecules that give and receive the electron. They generally do *not* match in the molecules' ground states, but a conformational fluctuation can momentarily bring them into agreement, again analogously to the situation in FRET (Section 14.3.2). J. Hopfield called this mechanism **thermally activated tunneling**.

Even at zero temperature, however, there is some probability for the molecules to be found in a favorable conformation, due to a second kind of quantum tunneling: The nuclei themselves have nonzero probability to be found in a region that is energetically forbidden. Hopfield found that he could explain the full curve in Figure 16.6 as a combination of nuclear tunneling at the lowest temperatures and thermal activation at higher temperature. In the system studied by DeVault and Chance, the crossover between regimes occurs at a temperature well below those experienced by living organisms. However, the success of the model throughout the temperature range studied strengthened the case that quantum tunneling is also responsible for electron transfer at ordinary temperatures. Other tunneling processes, even involving protons, have been documented in biomolecules and appear to play a role in their function.

 $^{^{12}}$ This result is analogous to what we found for FRET in Equation 14.16 (page 420).

Problems 455

THE BIG PICTURE

At first, photon tunneling appears paradoxical: If a photon does manage to escape the high-index region, what prevents it from traveling arbitrarily far? After all, the "forbidden" region is just another transparent medium, or even vacuum. What, then, explains the experimentally observed exponential decrease in probability density? Section 16.1 suggested that the *cancellation* we expect in the ray-optics approximation would be *incomplete* if the path length in the forbidden region is short. Then it is less surprising to find that deeper in the less dense medium the cancellation is more perfect, and a similar remark for electron tunneling. The derivations in Sections 16.2.2 and 16.3.2 used a mathematical trick to exchange the cancellation of many real phases for an overall imaginary phase. Similar overall behavior emerged in each context because an imaginary phase factor is the same as exponential suppression.

FURTHER READING

Semipopular: Al-Khalili & McFadden, 2014.

Intermediate:

Complex integrals: Cahill, 2013; Arfken & Weber, 1995.

Saddle-point approximation: Mathews & Walker, 1970; Arfken & Weber, 1995; Appel, 2007.

Evanescent wave calculated via classical wave theory: Zangwill, 2013, pp. 595–598. Arrhenius behavior of thermally activated reactions: Bialek, 2012, §§4.1, A.5. Electron and proton tunneling in biomolecules: Bialek, 2012, §4.1.

Technical:

Multilayer films with precise thickness: Blodgett, 1935.

Historic: Devault, 1984; Hopfield, 1974; DeVault & Chance, 1966.

PROBLEMS

16.1 Energy transfer versus distance

- a. H. Bücher and coauthors investigated the distance dependence of FRET by fabricating thin, planar layers of donor fluorophores each a precise distance from a second, parallel plane layer containing acceptors. Each donor therefore sees a plane of potential acceptors at varying distances $\sqrt{r^2 + z^2}$, where z is the distance between the planes, and r is distance in the acceptor plane from the point directly below the donor. Find a simple expression for the z dependence of the total energy transfer probability per time. Then write a formula analogous to Equation 2.3 giving the rate of donor fluorescence in terms of z and a phenomenological constant.
- b. The table below gives the observed ratio of donor fluorescence at each of several distances z to the value at infinite distance. See if you can fit it to the function you found in (a) with some suitable choice of the one fit parameter.