

Amplified total internal reflection: theory, analysis, and demonstration of existence via FDTD

Keely J. Willis¹, John B. Schneider², Susan C. Hagness¹

¹*Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706*

²*Department of Electrical Engineering and Computer Science, Washington State University, Pullman, WA 99164*

kjwillis@wisc.edu

Abstract: The explanation of wave behavior upon total internal reflection from a gainy medium has defied consensus for 40 years. We examine this question using both the finite-difference time-domain (FDTD) method and theoretical analyses. FDTD simulations of a localized wave impinging on a gainy half space are based directly on Maxwell's equations and make no underlying assumptions. They reveal that amplification occurs upon total internal reflection from a gainy medium; conversely, amplification does not occur for incidence below the critical angle. Excellent agreement is obtained between the FDTD results and an analytical formulation that employs a new branch cut in the complex "propagation-constant" plane.

© 2008 Optical Society of America

OCIS codes: (260.6970) Total internal reflection; (260.2110) Electromagnetic optics

References and links

1. C. J. Koester, "Laser action by enhanced total internal reflection," *IEEE J. Quantum Electron.* **2**, 580–584 (1966).
2. M. Born and E. Wolf, *Principles of Optics* (Pergamon Press, 1964).
3. G. N. Romanov and S. S. Shakhidzhanov, "Amplification of electromagnetic field in total internal reflection from a region of inverted population," *JETP* **16**, 298–301 (1972).
4. S. A. Lebedev, V. M. Volkov, and B. Ya. Kogan, "Value of the gain for light internally reflected from a medium with inverted population," *Opt. Spectrosc.* **35**, 976–977 (1973).
5. P. R. Callary and C. K. Carniglia, "Internal reflection from an amplifying layer," *J. Opt. Soc. Am.* **66**, 775–779 (1976).
6. R. F. Cybulski and C. K. Carniglia, "Internal reflection from an exponential amplifying region," *J. Opt. Soc. Am.* **67**, 1620–1627 (1977).
7. C.-W. Lee, K. Kim, J. Noh, and W. Jhe, "Quantum theory of amplified total internal reflection due to evanescent-mode coupling," *Phys. Rev. A* **62**, 053805 (2000).
8. J. Fan, A. Dogariu, and L. J. Wang, "Amplified total internal reflection," *Opt. Express* **11**, 299–308 (2003).
9. A. Taflov and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 2nd ed. (Artech House, 2000).
10. Z. Wang, Z. Zhang, Z. Xu, and Q. Lin, "Space-time profiles of an ultrashort pulsed Gaussian beam," *IEEE J. Quantum Electron.* **33**, 566–573 (1997).
11. A. Ishimaru, *Electromagnetic Wave Propagation, Radiation, and Scattering* (Prentice Hall, 1991).
12. D. J. Robinson and J. B. Schneider, "On the use of the geometric mean in FDTD near-to-far-field transformations" to appear in *IEEE Trans. Antennas Propag.* (2007).
13. J. B. Schneider, "Plane waves in FDTD simulations and a nearly perfect total-field/scattered-field boundary" *IEEE Trans. Antennas Propag.* **52**, 3280–3287 (2004).

1. Introduction

The physical description and even the existence of evanescent gain, or wave amplification that occurs within a passive medium upon total internal reflection (TIR) from a gainy medium, has been the subject of controversy for over 40 years. It has long been observed that an index-guided propagating mode in a dielectric optical waveguide with lossless core and gainy cladding will experience growth [1]. Consider instead the related scenario involving a wave propagating towards a dielectric interface with incident angle above the critical angle θ_c . Conventional optical theory predicts the reflection coefficient magnitude $\Gamma < 1$ for TIR regardless of whether the medium on the other side of the interface is gainy or lossy [2]. Thus, conventional theory predicts that propagating modes in a waveguide with gainy cladding and lossless core will experience loss (barring some other gain mechanism such as reflections from the outer edge of the cladding) in apparent conflict with the experimentally observed amplification of such modes. In this paper we show that, contrary to conventional theory, fields incident upon a gainy medium beyond the critical angle have reflection coefficients with magnitude greater than unity.

As shown in Fig. 1, we consider a region composed of two dielectric half spaces, where the dielectric interface is constant in x . We label the media filling these half spaces as medium 1 and medium 2, where medium 1 is to the left of the interface and relates to the lossless core of the optical waveguide, and medium 2 is to the right and relates to the waveguide's cladding. To extract information purely concerning reflection from the interface between the two media and avoid contributions from back reflections within medium 2, we eliminate the back interface corresponding to the outside of the cladding so that medium 2 extends to infinity in the x direction. Both media are nonmagnetic, with dielectric constants ϵ_{r1} and ϵ_{r2} and conductivities σ_1 and σ_2 . Consistent with a positive index-step optical waveguide, and to allow for TIR, we require $\epsilon_{r2} < \epsilon_{r1}$. Medium 1 is lossless, so $\sigma_1 = 0$. Medium 2 is either *lossy* if $\sigma_2 > 0$ or *gainy* if $\sigma_2 < 0$. The conductivity σ_2 is taken to be a constant independent of frequency.

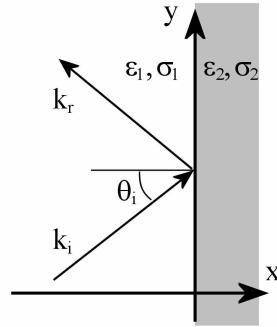


Fig. 1. System under consideration: dielectric half spaces. Medium 1 corresponds to the lossless core of a positive index-step optical waveguide, and medium 2 corresponds to either the gainy or lossy cladding.

Assuming $\exp(j\omega t)$ temporal dependence, the spatial dependence for the incident and reflected fields are given by $\exp(-j\mathbf{k}_i \cdot \mathbf{r})$ and $\exp(-j\mathbf{k}_r \cdot \mathbf{r})$, respectively, where $\mathbf{k}_i = k_{x1}\hat{\mathbf{a}}_x + k_{y1}\hat{\mathbf{a}}_y = k_1(\cos\theta_i\hat{\mathbf{a}}_x + \sin\theta_i\hat{\mathbf{a}}_y)$. The reflected-field wave vector differs from that of the incident field only in the sign of the x -component so that $|\mathbf{k}_i| = |\mathbf{k}_r| = k_1 = \omega(\epsilon_{r1}\epsilon_0\mu_0)^{1/2}$. Spatial dependence in the second medium is given by $\exp((\mathbf{g} - j\mathbf{k}_t) \cdot \mathbf{r})$. The orientation of the gain vector \mathbf{g} corresponds to the direction of increasing amplitude. Owing to continuity of the tangential

fields across the interface, $k_{y2} = k_{y1}$ and \mathbf{g} has only an x component.

The s-polarized Fresnel reflection coefficient can now be expressed as

$$R = \frac{k_1 \cos \theta_i - (k_{x2} + jg_x)}{k_1 \cos \theta_i + (k_{x2} + jg_x)}, \quad (1)$$

$$= \Gamma \exp(j\phi).$$

Alternatively, using the notation of [2], one can write the reflection coefficient as

$$R = \frac{n_1 \cos \theta_i - (u_2 + jv_2)}{n_1 \cos \theta_i + (u_2 + jv_2)} \quad (2)$$

where

$$2u_2^2 = n_2^2(1 - \kappa_2^2) - n_1^2 \sin^2 \theta_i + \sqrt{[n_2^2(1 - \kappa_2^2) - n_1^2 \sin^2 \theta_i]^2 + 4n_2^4 \kappa_2^2}, \quad (3)$$

$$2v_2^2 = -n_2^2(1 - \kappa_2^2) + n_1^2 \sin^2 \theta_i + \sqrt{[n_2^2(1 - \kappa_2^2) - n_1^2 \sin^2 \theta_i]^2 + 4n_2^4 \kappa_2^2}. \quad (4)$$

Here n_1 is the index of refraction of medium 1, and κ_2 is related to the imaginary part of the complex index of refraction of medium 2, \hat{n}_2 , as follows:

$$\hat{n}_2 = n_2(1 + j\kappa_2) = \sqrt{\epsilon_{r2} - j \frac{\sigma_2}{\omega \epsilon_0}}. \quad (5)$$

A change in the sign of σ_2 changes only the sign of κ_2 in Eq. (5). Because κ_2 is squared everywhere it appears in Eqs. (3)–(4), only its magnitude affects the value of R so that the Fresnel reflection coefficient as given by Eqs. (2)–(5) is not affected by the sign of σ_2 . This conventional application of optical theory shows evanescent gain not to occur without some new interpretation of the theory.

As a result, early observations of gain in the propagating modes of an optical waveguide with lossless core and gainy cladding incited numerous explanations; a small sampling of these are described here. Romanov and Shakhidzhanov (R&S) posited that gain should be possible upon TIR from a gainy medium and forbidden for incidence below the critical angle [3]. Shortly after these results were published, Lebedev, Volkov and Kogan (LV&K) produced experimental values for the reflection coefficient that conflicted with the predictions of the R&S theory [4]. Several years later, Callary and Carniglia presented an alternate theory which predicted gain upon reflection for incidence below the critical angle, as well as above [5]. Soon after, Cybulski and Carniglia proposed a theory to explain the LV&K result, but did not find agreement [6].

The question of the mechanism of evanescent gain has been the subject of a more recent surge of interest. Citing the lack of consensus in the scientific community and the broad range of predicted values for gain upon reflection given in the literature, Lee, Kim, Noh and Jhe asserted that bulk material properties cannot be used to describe evanescent gain [7]. The Lee *et al.* work insisted that a full quantum mechanical approach must be taken to accurately describe this phenomenon, and it did not make the connection to classical theory. More recently Fan, Dogariu and Wang reexamined the theory as given by R&S, and used numerical methods to predict the Goos-Hänchen shift that occurs upon reflection from gainy and lossy media [8]. Overall, past research has generally agreed that gain can occur upon reflection from a gainy medium and that the amount of gain depends on the incident angle; however, a vast disparity exists in predictions for the circumstances under which gain will occur, and the amount of gain that will occur upon reflection.

Because u_2 and v_2 are squared in Eqs. (3) and (4), their definition contains an inherent sign ambiguity. The presence of this sign choice produces two solutions to Maxwell's equations

which are *mathematically* valid. Which solution is *physically* valid? The sign choice affects the resulting Fresnel reflection coefficient magnitude Γ . Conventionally the choice of signs for u_2 and v_2 correspond to taking k_{x2} to be positive and g_x to be negative for $\sigma_2 \neq 0$. This yields phase propagation away from the interface and exponential amplitude decay away from the interface in medium 2.

These sign choices stem from the conventional branch cut in the complex plane of system parameters, which assumes that the reflection coefficient varies continuously with incident angle. Is this assumed branch cut correctly defined? Historically, neither theoretical analysis nor experiment has provided a definitive answer, as proven by the enduring lack of consensus within the scientific community on this point. Recent advances in computational electromagnetics techniques and an increase in computing power permit an alternate analysis using the finite-difference time-domain (FDTD) method [9]. A full-wave time-domain Maxwell's equations solver, FDTD allows for precise control of system parameters impossible in experiment and makes none of the underlying assumptions that are required in theoretical analyses. FDTD is ideally suited to examine this system and answer the question of sign choice.

In this paper we present a new examination of the existence and properties of evanescent gain, using first-principles FDTD analysis. First we model the interaction of a localized wave (a Gaussian beam) with the gainy half space. Using the FDTD-computed reflection coefficients as a benchmark, we define a new branch cut in the complex plane of system parameters. The choice of this branch cut is made specifically so that the analytical description of the reflection coefficient for the localized wave, which relies on a branch cut definition, achieves agreement with the FDTD results. We believe that for a medium with constant conductivity the outcome of this first-principles investigation is a conclusive description of the circumstances under which evanescent gain will occur and the elimination of ambiguities in the Fresnel equations.

2. FDTD analysis of Gaussian beam reflection

To examine the interaction of a localized wave with this dielectric interface we use a 2D FDTD grid (defined as a TM_z grid with non-zero E_z , H_x , and H_y field components). The computational domain is shown in Fig. 2. For these simulations $\epsilon_{r1} = 4$ and $\epsilon_{r2} = 2$. Both media are non-magnetic, as before. Again, medium 1 is taken to be lossless, and medium 2 has conductivity $\sigma_2 = \pm 500 \text{ S/m}$, where we take the upper sign for a lossy medium and the lower sign for a gainy medium. The incident field is a Gaussian beam that is pulsed in time with a Gaussian envelope and carrier frequency $f_0 = 3.33 \times 10^{14} \text{ Hz}$. In medium 1 there are 20 samples per wavelength at the center frequency, and the Courant number is $0.95/\sqrt{2}$ in free space. We use the conventional definition of the critical angle, so that here $\theta_c = 45^\circ$.

To model the interaction of the incident wave with a true half space and thereby create reflections from only the single interface, we terminate the grid boundaries with uniaxial perfectly matched layer (PML) absorbing boundary conditions for general (including conductive) media [9]. The PML is designed to absorb outward-propagating waves that reach the grid boundary. The performance of PML in terminating lossy media has been documented extensively in the literature. We validated the performance of the PML in providing essentially a reflection-free termination of the gainy medium by comparing the reflections from the PML with those from a perfect electric conductor (PEC). We found that the PML reflection at normal incidence was 140 dB below that from the PEC. In the case of evanescent decay, we require that the interface be defined far enough from the PML in medium 2 that the full decaying field be represented within the FDTD grid; thus we require at least a few wavelengths of dielectric in medium 2 between the interface and right-hand PML. As the incident radiation approaches and passes through the critical angle, any outgoing radiation in medium 2 is absorbed by the PML and does not reflect back into the main part of the grid to interact with the interface again; this

is essential to properly represent the semi-infinite half space.

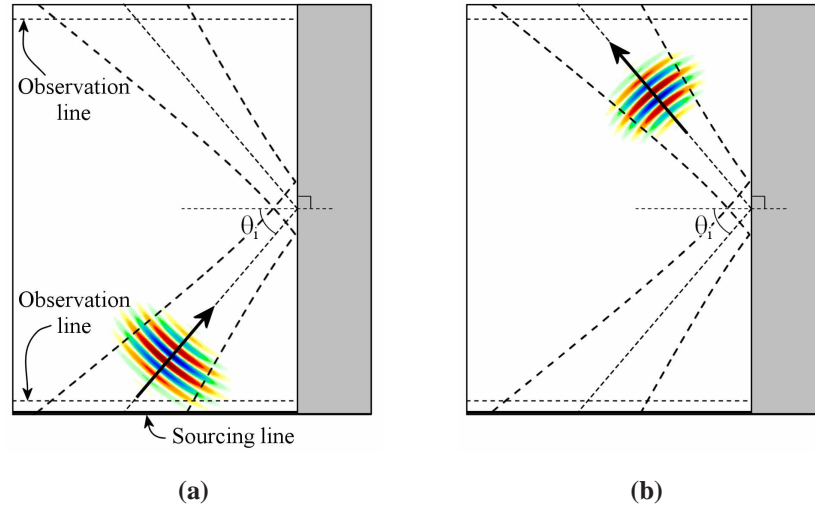


Fig. 2. Snapshots of a pulsed-in-time Gaussian beam launched into the FDTD grid, where red indicates positive values of E_z and blue indicates negative values of E_z . (a) shows the incident pulse, and (b) shows the pulse after reflection from the dielectric interface.

2.1. Incident wave source condition

To launch the localized wave in the FDTD grid, we specify E_z fields along a sourcing line (the bottom-most row of grid cells, as shown in Fig. 2) at each FDTD time step. The sourcing line forms an angle θ_i with the desired beam axis as shown in Fig. 3. The source field values are pre-calculated using an analytical description of a Gaussian beam modulated with a Gaussian pulse [10]. By controlling the amplitude and phase of the fields along the sourcing line, we are able to launch a pulse with a chosen direction of propagation into the FDTD grid. Fig. 2(a) shows a snapshot in time of the pulse after it has been launched from the sourcing line. It propagates up and to the right to intersect the interface with incident angle θ_i with respect to the interface normal. As long as we respect the paraxial approximation inherent to the analytical description of the Gaussian beam, and use a reasonable grid resolution in the FDTD simulations, we have great latitude in manipulating the Gaussian beam parameters. Here the beam waist W_0 is modified through its association with the Rayleigh diffraction length, and the incident angle of the pulse is adjusted to sweep through the critical angle for TIR.

2.2. Reflection coefficient calculation

Figure 2 shows two observation lines in medium 1 along which the E_z and H_x components of the incident and reflected waves are recorded. A discrete Fourier transform is used to obtain the field phasors at the carrier frequency f_0 from the time-domain fields. The observation lines run through the centers of the Yee grid cells, so that the E_z components are located on the observation lines and the H_x components are positioned a half grid cell above and below the line.

After completion of the time-stepping loop, one set of field phasors (either E_z or H_x) is multiplied by a phase factor so that all field components use a common temporal reference point. This has the effect of removing the stagger in time between the electric and magnetic

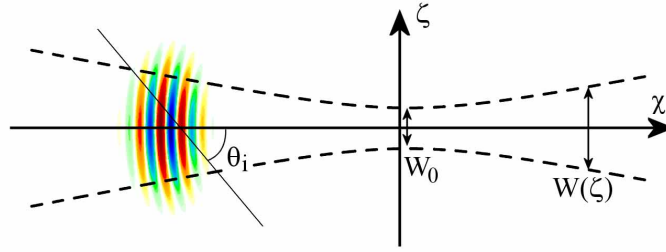


Fig. 3. Gaussian beam time-modulated with a Gaussian pulse. The diagonal line where E_z is pre-calculated corresponds to the bottom-most row of E_z field components in the FDTD grid.

field that is inherent in the FDTD method. Additionally, the geometric mean is taken of the magnetic fields to either side of the electric fields on the observation line. This serves to spatially collocate the electric and magnetic fields [12]. These unstaggered and collocated fields are used to calculate the y-directed power density at each point along the observation lines. The power density values are then summed to compute the incident and reflected y-directed power [W/m in 2D] that passes across each line. The square root of the ratio of the reflected power to the incident power yields the magnitude of the reflection coefficient.

2.3. FDTD results for the Gaussian beam

In Sec. 2.1 we defined the incident angle of the Gaussian beam as the angle of the beam axis with respect to the interface normal. In reality the spectral content of any single beam results in incidence at a range of angles, and the chosen single “ θ_i ” in our calculations is an approximation for the weighted average of the full spectrum of incident angles. (For true incidence with a single θ_i we would use plane wave analysis.) Each angular component of the incident field has a reflection coefficient that depends on the component angle with respect to the interface normal; the sum of all such field components produces the beam reflection coefficient.

The angular spread of a Gaussian beam is inversely related to its spatial width with a plane wave serving as the limit of an infinitely wide beam with a single incident angle. The various curves with data markers in Fig. 4 show the FDTD-computed reflection coefficient magnitude at the carrier frequency as a function of incident angle for several beam widths. Clearly the incident field angular spread strongly affects the resulting beam reflection coefficient near the critical angle. We now have some explanation for the disparity in reported experimental results: gain upon reflection is strongly dependent on the angular spread of the incident beam, especially near the critical angle.

As will be discussed in more detail in the next section, the solid black line in Fig. 4 is the analytical Fresnel reflection coefficient where we have taken k_{x2} to be negative for incident angles greater than θ_c . Figure 4 shows that as the width of the beam waist increases, the FDTD beam results approach the analytical plane wave results obtained using this non-conventional sign choice. The mathematical justification for this sign choice is discussed below.

3. Theoretical analysis: a new branch cut in the complex plane

As stated previously, the spatial dependence in medium 2 is given by

$$\exp((\mathbf{g} - j\mathbf{k}_t) \cdot \mathbf{r}) = \exp(-jk_{y1}y - [-g_x + jk_{x2}]x) \quad (6)$$

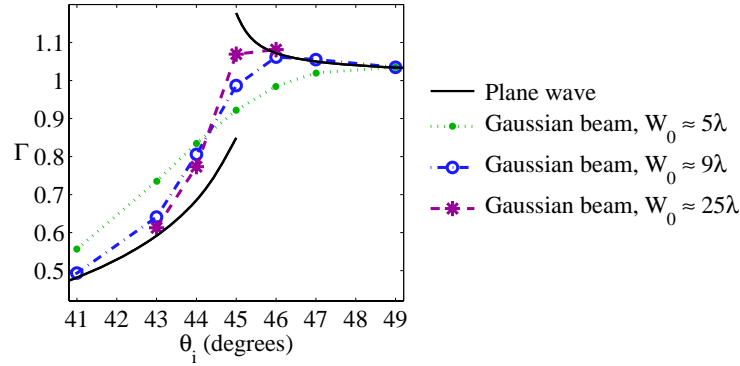


Fig. 4. Comparison of three FDTD-computed beam reflection coefficient magnitudes for a gainy half space with a critical angle of 45° . The three different beam waists W_0 are given in terms of λ , the wavelength at the carrier frequency in medium 1. The plane wave reference curve is computed using Fresnel theory, taking k_{x2} to be negative for $\theta_i > \theta_c$.

Importantly, if g_x and k_{x2} are known unambiguously, then from Eq. (1) the reflection coefficient is also known unambiguously.

Applying the wave equation to any field component in medium 2 yields the dispersion equation

$$(-g_x + jk_{x2})^2 - k_{y1}^2 - j\omega\mu_0(\sigma_2 + j\omega\epsilon_2) = 0. \quad (7)$$

Solving for $(-g_x + jk_{x2})$ yields

$$-g_x + jk_{x2} = k_0 \left(\epsilon_{r1} \sin^2 \theta_i - \epsilon_{r2} + j \frac{\sigma_2}{\omega\epsilon_0} \right)^{1/2} \quad (8)$$

where $k_0 = \omega\sqrt{\mu_0\epsilon_0}$. Let the complex quantity ξ represent the term in parentheses so that

$$\xi = \xi_r + j\xi_i = (\epsilon_{r1} \sin^2 \theta_i - \epsilon_{r2}) + j \left(\frac{\sigma_2}{\omega\epsilon_0} \right). \quad (9)$$

In Eq. (9) the real and imaginary parts of ξ are grouped in parentheses. The real part ξ_r is only a function of the incident angle θ_i for given material parameters, while the imaginary part ξ_i is only a function of the conductivity σ_2 for a given frequency. We can now write

$$-g_x + jk_{x2} = k_0 \xi^{1/2}. \quad (10)$$

Since k_0 is real, to uniquely determine g_x and k_{x2} our task is reduced to determining the proper branch cut for $\xi^{1/2}$. (Thus, rather than determining which of the four possible combinations of signs for g_x and k_{x2} is correct, we must determine one branch cut.)

Regardless of the conductivity, we define the critical angle θ_c to be

$$\theta_c = \sin^{-1} \left(\sqrt{\frac{\epsilon_{r2}}{\epsilon_{r1}}} \right). \quad (11)$$

At the critical angle the real part of ξ is zero. We will say that the field is *evanescent* in medium 2 if $\theta_i > \theta_c$ ($\xi_r > 0$). Conversely, the field is said to be *propagating* in medium 2 if $\theta_i < \theta_c$ ($\xi_r < 0$). Again we note that we adopt the terms evanescent or propagating independent of the conductivity (since only the incident angle affects the real part of ξ).

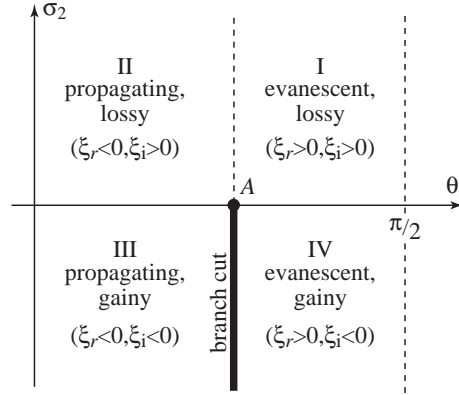


Fig. 5. The (θ_i, σ_2) -plane. Each point in the plane maps to a unique value of ξ as given by Eq. (9). The point A corresponds to $(\theta_i = \theta_c, \sigma_2 = 0)$, i.e., $\xi = 0$ where the field is neither propagating nor decaying in the x direction. The incident angle θ_i is between zero and $\pi/2$. The conductivity σ_2 can take on any real value. The four quadrants are distinguished by whether they represent a propagating or evanescent field in either a lossy or a gainy medium. These quadrants also correspond to the different signs for the real and imaginary parts of ξ .

At this point it is instructive to visualize a two-dimensional plane in which the independent and orthogonal coordinates are the incident angle and the conductivity in medium 2, i.e., the “ (θ_i, σ_2) -plane,” which is shown in Fig. 5. We consider incident angles θ_i between 0 and $\pi/2$ while the conductivity can take on any real value.

The point A in Fig. 5 corresponds to $(\theta_i = \theta_c, \sigma_2 = 0)$ —a wave incident on a lossless medium at the critical angle. At this point $\xi = 0$ so that the field in medium 2 would be neither propagating nor decaying in the x direction. As will be discussed in a moment, the branch cut (which is really pertinent to the ξ -plane and not the (θ_i, σ_2) -plane) will be defined as the line segment $(\theta_i = \theta_c, \sigma_2 < 0)$ which corresponds to the heavy line drawn below the point A.

The ξ -plane is shown on the left side of Fig. 6. The corresponding quadrants from the (θ_i, σ_2) -plane carry over directly as shown by the quadrant labels. The branch cut corresponds to the negative imaginary axis. Given the branch cut shown in Fig. 6, an arbitrary point in the ξ -plane can be written as

$$\xi = |\xi| e^{j\theta_\xi} \quad \text{where} \quad -\frac{\pi}{2} \leq \theta_\xi < \frac{3\pi}{2}. \quad (12)$$

Returning to g_x and k_{x2} , we have

$$-g_x + jk_{x2} = k_0 \xi^{1/2} = k_0 (\xi_r^2 + \xi_i^2)^{1/4} e^{j\theta_\xi/2}. \quad (13)$$

Thus one can express g_x and k_{x2} as

$$g_x = -k_0 (\xi_r^2 + \xi_i^2)^{1/4} \cos(\theta_\xi/2), \quad (14)$$

$$k_{x2} = k_0 (\xi_r^2 + \xi_i^2)^{1/4} \sin(\theta_\xi/2). \quad (15)$$

The mapping of the ξ -plane to the (g_x, k_{x2}) -plane is shown in Fig. 6.

The quadrant of a point is dictated by the value of θ_ξ while the branch cut dictates the valid range of θ_ξ . Given this, the gain g_x and phase constant k_{x2} take on the following values:

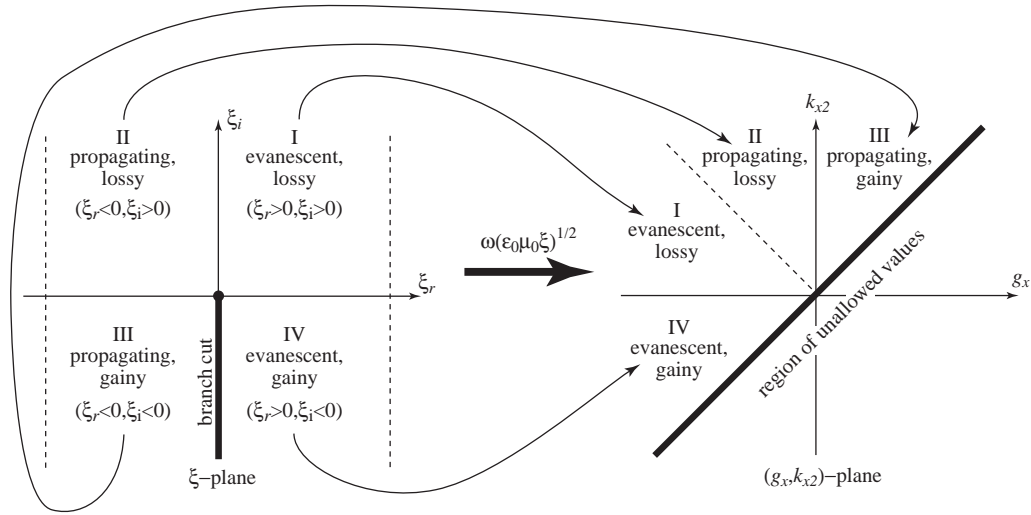


Fig. 6. The mapping of the ξ -plane to the (g_x, k_{x2}) -plane. Points in the (g_x, k_{x2}) -plane are obtained from the square root of points in the ξ -plane scaled by $\omega\sqrt{\mu_0\epsilon_0}$. In the (g_x, k_{x2}) -plane points below and to the right of the heavy diagonal line are not realizable. The sector labels shown in the (g_x, k_{x2}) -plane are taken from the corresponding quadrant labels in the ξ -plane.

- If $-\pi/2 \leq \theta_\xi < 0$ (corresponding to incidence greater than the critical angle and a gainy second medium), then $g_x < 0$ and $k_{x2} < 0$. Thus the amplitude of the field decays in the positive x direction and the phase propagation is in the negative x direction.
- If $0 \leq \theta_\xi < \pi$ (corresponding to an arbitrary angle of incidence and a lossy second medium), then $g_x < 0$ and $k_{x2} > 0$. Thus, as was true in the previous case, the amplitude of the field decays in the positive x direction but now phase propagation is in the positive x direction.
- If $\pi \leq \theta_\xi < 3\pi/2$ (corresponding to incidence less than the critical angle and a gainy second medium), then $g_x > 0$ and $k_{x2} > 0$. Therefore the amplitude increases in the positive x direction and phase also propagates in the positive x direction.

Referring back to Fig. 5, the dividing line between quadrants I and IV is given by $(\theta_c < \theta_i \leq \pi/2, \sigma_2 = 0)$. For these values of θ_i and σ_2 the field in medium 2 is purely evanescent and there is no loss or gain. Along this line k_{x2} is zero while g_x is non-zero and it varies continuously to either side. This line essentially corresponds to the branch cut that has been advocated in the past (which yields $0 \leq \theta_\xi < 2\pi$). Were one to use this line as the branch cut there would be a discontinuity in g_x as a function of conductivity for a fixed angle.

No matter where the branch cut is placed, as one varies either the conductivity or the incident angle a discontinuity will and must exist in either g_x or k_{x2} or in both. For the branch cut we propose, there is a discontinuity in both g_x and k_{x2} (a change in sign of both quantities) as one passes through the critical angle for negative conductivity. Such a discontinuity is not unprecedented as the Goos-Hänchen shift is absent for angles less than the critical angle but present for any angle greater than the critical angle [11]. The shift is either there or it is not. It does not vary continuously and the discontinuity occurs at the critical angle. In the same spirit, and consistent with the results obtained from the FDTD simulations, we have kept the critical angle as the branch cut even when gain is present.

Note that region IV is the only region where *phase propagation is directed back toward the interface*. The gain vector is also directed toward the interface. This produces a reflection coefficient with a magnitude greater than unity (this is only true for region IV). Thus, a slight change in incident angle, i.e., a change that transitions from the propagating to the evanescent side of the branch cut, will result in the magnitude of the reflection coefficient changing from a value less than unity to one greater than unity. This “switch flipping” will happen discontinuously as shown for the plane wave curve in Fig. 4. This discontinuity is consistent with all the FDTD simulations we have performed.

3.1. Theoretical analysis of Gaussian beam reflection

The pulsed incident wave can be decomposed into its frequency-domain components and, for any given frequency, the corresponding Gaussian beam can be represented as a continuous spectrum of plane waves as depicted in Fig. 7. The $\chi\zeta$ -coordinate system is such that the central beam axis is aligned with the χ axis thus forming the angle θ_i between the x and χ axes.

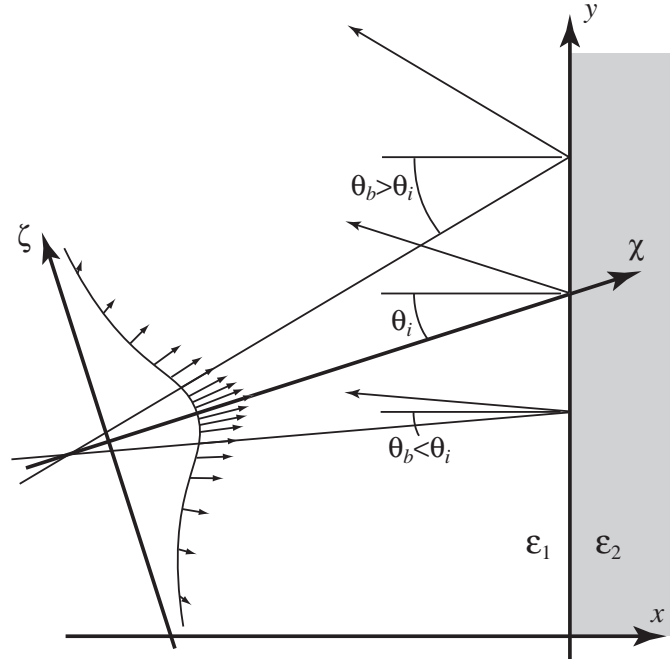


Fig. 7. Decomposition of an incident Gaussian beam to a spectrum of plane waves. The reflected beam is similar to the incident beam except it is weighted by the reflection coefficient and propagates to the left.

The incident field is given by

$$E_z^i(\chi, \zeta) = \frac{1}{2\pi} \int A(k_\zeta) e^{-jk_\zeta \zeta - jk_\chi \chi} dk_\zeta, \quad (16)$$

where the amplitude of the angular spectrum is

$$A(k_\zeta) = \pi W_0^2 e^{-W_0^2 k_\zeta^2 / 4} \quad (17)$$

and k_ζ and k_χ represent the wave numbers in the ζ and χ directions, respectively. A component

of the beam not propagating along the beam axis forms an incident angle θ_b with the x axis. This angle can be greater than or less than the nominal incident angle θ_i .

The reflected beam is obtained from the incident-beam spectrum but a sign change accounts for the fact that the reflected field propagates in the $-\chi$ direction and the reflection coefficient weights the angular spectrum. The reflected field is thus given by

$$E_z^r(\chi, \zeta) = \frac{1}{2\pi} \int R(k_\zeta, \theta_i) e^{-jk_\zeta \zeta + jk_\chi \chi} dk_\zeta. \quad (18)$$

Because of the geometry of the FDTD simulation, only wave numbers corresponding to propagating waves with “beam angle” θ_b between 0° and 90° are considered. Although the reflection coefficient changes depending on the branch cut that is used, the form of Eq. (18) is independent of the branch cut. (We note that in practice, unlike the depiction in Fig. 7, our calculations assume the origin of the $\chi\zeta$ -coordinate system is collocated with the origin of the xy -coordinate system. Since we are ultimately interested in power and not the fields, the choice of the origin is immaterial.)

From the incident and reflected fields, one can obtain the incident and reflected power. Taking the square root of the ratio of these two yields the beam reflection coefficient.

3.2. Comparison of analytical and FDTD results for the Gaussian beam

The results of the FDTD beam reflection coefficient calculations for both the gainy and lossy cases are shown in Fig. 8. Also shown are the analytically calculated beam reflection coefficients when using the new branch cut. The Fresnel plane wave reflection coefficients calculated assuming either the new branch cut or the conventional branch cut are also plotted. The clear agreement between our FDTD result, which depends purely on Maxwell’s equations, and the analytical result that employs the new branch cut, gives strong support for the use of this branch cut.

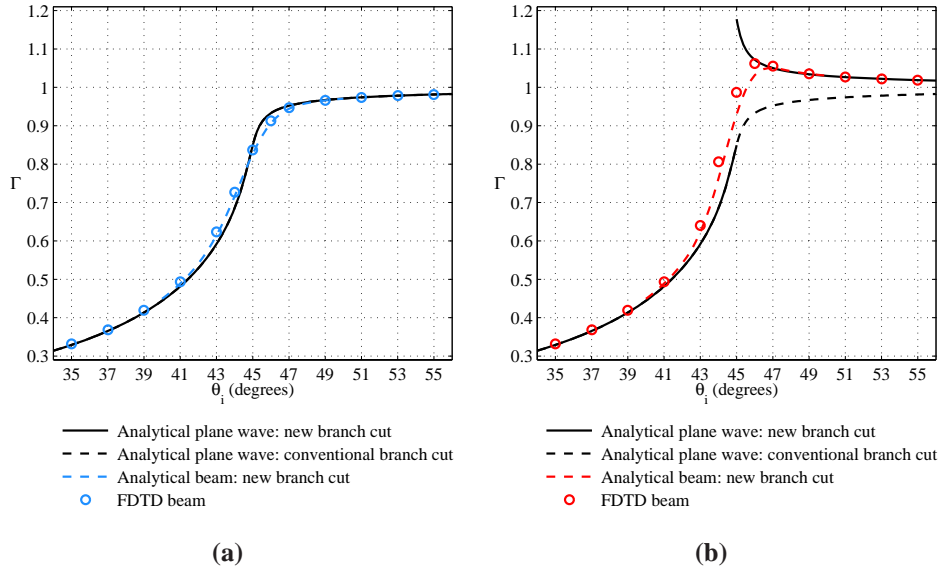


Fig. 8. Comparison of reflection coefficient magnitudes computed for the incident beam using the analytical and FDTD formulations. (a) shows results for reflection from the lossy half space, and (b) shows results for reflection from the gainy half space.

The small difference between the FDTD and analytical data can be attributed to the numerical approximations inherent to the FDTD formulation, as well as the fact that the analytical solution has no curvature in the phase fronts of the Gaussian beam. The phase front curvature affects the incident-angle spectrum which, in turn, affects the value of Γ .

To resolve fine angle-dependent detail in the reflection coefficient calculation we employ the analytic field propagation total-field/scattered-field (AFP-TFSF) technique, a recent advance in FDTD methods which allows for high-accuracy modeling of plane wave interaction with the dielectric interface [13]. Plane wave illumination allows for incoming radiation with a single incident angle, permitting further verification of the precise location of the branch cut in Fig. 5. Further, the analytical field descriptions employed in the AFP-TFSF technique explicitly demonstrate the agreement of Maxwell's equations with the backwards phase propagation in medium 2 described by the new branch cut. A detailed discussion of the AFP-TFSF technique and plane wave simulations is beyond the scope of this paper.

4. Conclusion

We have conducted rigorous FDTD simulations of the interaction of a pulsed Gaussian beam with a gainy half space both above and below the critical angle. We compared the FDTD-computed reflection coefficient for the beam with the analytically derived reflection coefficient, where the analytical case assumed a new, non-conventional branch cut in the complex plane defined by the critical angle. The excellent agreement that was observed between these two sets of data supports the validity of this branch cut for a medium with constant conductivity.

This new branch cut creates a clear connection between system parameters and sign choice in the Fresnel equations. Once a branch cut is fixed there is no sign ambiguity associated with the square root. Through pulsed beam analysis, FDTD has confirmed the validity of the new branch cut. This research has shown that evanescent gain occurs for incidence above the critical angle and not for incidence below the critical angle. These results agree with the original theory presented by Romanov and Shakhidzhanov as well as the more recent reworking of theory by Fan, Dogariu and Wang.

Acknowledgment

The authors would like to thank Anthony E. Siegman of Stanford University for drawing their attention to this problem and for many interesting and enlightening discussions of theory. This work was supported by the National Science Foundation Presidential Early Career Award for Scientists and Engineers ECCS-9985004 and the Wisconsin Alumni Research Foundation.