A Time-Domain Analysis of Enhanced Total Internal Reflection Using the FDTD Method

Zhen Chen, John B. Schneider, *Fellow, IEEE*, Keely Willis, *Member, IEEE*, and Susan C. Hagness, *Fellow, IEEE*

***Abstract*—There is a long-standing debate surrounding whether or not *enhanced* total internal reflection (ETIR) is possible. ETIR implies that the magnitude of the reflection coefficient is greater than unity and is conjectured to be possible when a field is incident from a lossless material to a gainy material beyond the critical angle. In this letter, we examine this problem through finite-difference time-domain (FDTD) modeling. The two-dimensional simulations employ a Gaussian incident beam and make no *a priori* assumptions about the reflection coefficient. We consider illumination of gainy, lossless, and lossy materials. The Poynting vector is used to examine the flow of energy. For a gainy material, the magnitude of the reflection coefficient is found to be greater than unity, but there is a delay between when energy enters the gainy material and when the "excess" energy is reflected from the interface. Thus, given the Goos–Hänchen shift associated with total internal reflection, where the reflected beam is shifted relative to the incident beam (so that fields must travel in the gainy material before being reflected), the existence of ETIR appears not only to be plausible, but to be inevitable.**

***Index Terms*—Electromagnetic reflection, finite-difference time-domain (FDTD) methods, reflection coefficient.**

I. Introduction

AS DEPICTED in Fig. 1, total internal reflection (TIR) occurs when a field is incident beyond the critical angle from a material with a higher refractive index (the first medium) to one with a lower refractive index (the second medium). TIR has long been the subject of both research and industrial applications, e.g., [1]–[5]. It is an accepted fact that the magnitude of the reflection coefficient is less than unity [6] when the second medium is lossy and is unity when the second medium is lossless. On the other hand, there has been considerable debate whether or not the magnitude of the reflection coefficient can be greater than unity when the second medium is gainy, e.g., [7]–[10]. Since the first laboratory observation of the enhanced total internal reflection (ETIR) in a fiber with neodymium-doped glass cladding by Koester in 1966 [11], there have been numerous experimental and theoretical studies [7]–[10], [12]–[14] that attempted to explain this phenomenon, but none have succeeded in ending the debate.

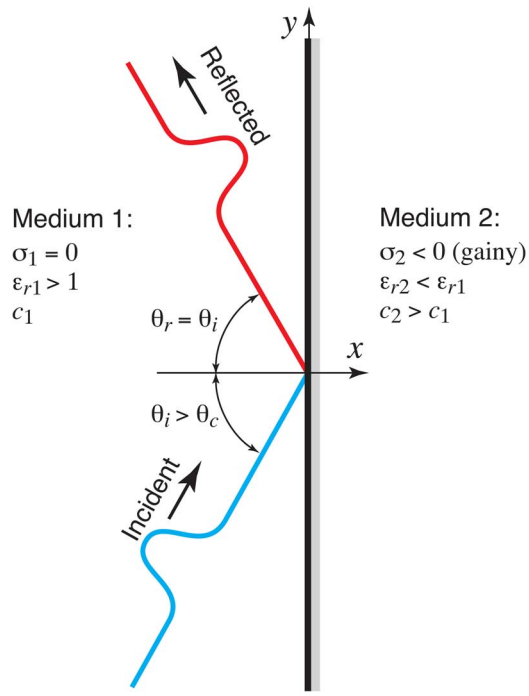

Fig. 1. Depiction of total internal reflection where the incident field is incident beyond the critical angle. The incident field is a pulsed (and modulated) Gaussian beam. (The Goos–Hänchen shift is not shown.)

One of the reasons for the persistent debate is that the analytic derivation of the reflection coefficient involves a square root that, in a pure mathematical sense, allows one to select either the positive or negative root and still obtain an equally valid solution. Unfortunately, there is no equivalent to the radiation condition to provide physical reasons to choose one root over another. If one views the reflection coefficient as being representative of what happens solely at the surface of an interface, that motivates selecting the root that yields a reflection coefficient with a magnitude less than unity. After all, how could one get more energy out of a system than was put in by merely reflecting a field from an interface? However, as we discuss in this letter, viewing the reflection coefficient as being merely dictated by the impedance mismatch at an interface is too simplistic.

Romanov and Shakhidzhanov [12] first derived a reflection coefficient with a magnitude greater than unity by choosing appropriate signs for the square roots associated with the Fresnel

Manuscript received November 27, 2013; revised January 06, 2014; accepted January 12, 2014. Date of publication January 14, 2014; date of current version February 12, 2014.

Z. Chen and J. B. Schneider are with the School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA 99164 USA (e-mail: zchen@eecs.wsu.edu; schneidj@eecs.wsu.edu).

K. Willis is with AWR Corporation, Mequon, WI 53092 USA.

S. C. Hagness is with the Department of Electrical and Computer Engineering, University of Wisconsin–Madison, Madison, WI 53706 USA.

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LAWP.2014.2300138

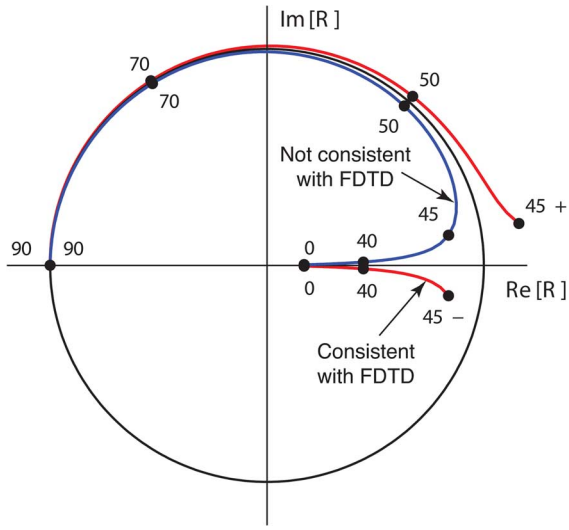

Fig. 2. Plot of the reflection coefficient R versus incident angle θ_i in the reflection-coefficient plane. The circle is the unit circle. The curve labeled "Consistent with FDTD" employs the branch cut that yields analytic results consistent with previous FDTD simulations [9]. Note the discontinuity in the reflection coefficient at the critical angle (45°). Furthermore, note that the magnitude of the reflection coefficient is greater than unity for incident angles beyond the critical angle. The curve labeled "Not consistent with FDTD" employs the branch cut advocated by some, e.g., [10]. In this case, there is no discontinuity, and the magnitude is always less than unity.

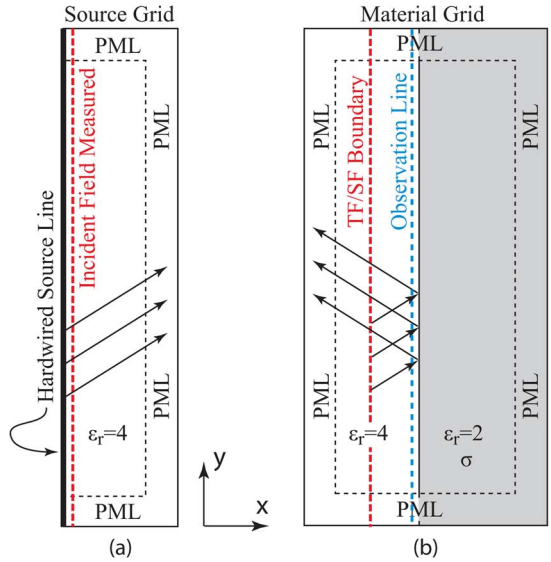

Fig. 3. Depiction of the two 2-D TM^z FDTD grids in the simulations. (a) Auxiliary grid that provides the incident field for the TF/SF boundary in the "material grid." (b) Material grid.

equations. Consistent with results obtained from finite-difference time-domain (FDTD) simulation where the reflection coefficient was found to have a magnitude greater than unity, Willis *et al.* [9] proposed an analytic solution that specified a particular location for the branch cut in the complex plane associated with the Fresnel equation square roots. This branch cut was consistent with the earlier work of Fan *et al.* [8], who also extensively investigated the Goos–Hänchen shift [15]. Fig. 2 demonstrates the consequences of selecting different branch cuts for the Fresnel equations. The complex reflection-coefficient plane shows the reflection coefficient versus incident angle when $\varepsilon_{\mathrm{r}1} = 4$, $\varepsilon_{\mathrm{r}2} = 2$, $\sigma_2 = -500$ S/m, and the frequency is 3.33×10^{14} Hz (see the figure caption for further details).

In this letter, we investigate ETIR via the time-domain behavior of Poynting vectors. In Section II, we describe the FDTD simulations that were performed. In Section III, we provide results that show the total amount of energy flow across the interface when the second medium is either gainy, lossless, or lossy. Snapshots of the Poynting vector near the interface are provided to give an intuitive interpretation of the ways in which the different media behave. Moreover, a trace of the location of the maximum energy in the pulsed beam is presented to demonstrate the Goos–Hänchen shift observed in the FDTD simulations. This shift appears to be a key component of ETIR.

II. Analysis and Numerical Approaches

The analysis was conducted using the "material grid" depicted in Fig. 3(b), which is a 2-D TM^z FDTD grid consisting of two homogeneous nonmagnetic media where the first medium has $\varepsilon_{\mathrm{r}1} = 4$ and the second has $\varepsilon_{\mathrm{r}2} = 2$ (thus the critical angle is 45°). The interface is the straight line parallel to the y-axis. The first medium is lossless, while the second is either gainy, lossless, or lossy with corresponding conductivities σ_2 of -500, 0, or 500 S/m, respectively (if not otherwise specified). Simulations of the gainy material were done using both a dispersive implementation and one using constant material parameters. At the frequency of interest, where the material values were equal, there was no difference in the results. The results shown here used the constant parameter model. To model a semi-infinite half-space, we terminate the grid by a convolutional perfectly match layer (CPML) [16]. We launch a localized Gaussian beam in the FDTD grid using an auxiliary source grid as shown in Fig. 3(a). The beam was launched into this auxiliary grid by using a hard-wired electric-field source along the left edge of the grid. This grid was simply used to obtain the $\mathbf{E}$ and $\mathbf{H}$ of the incident beam with the same dielectric constant as the first medium in the material grid [these fields were obtained at the line labeled "Incident Field Measured" in Fig. 3(a)]. Using this $\mathbf{E}$ and $\mathbf{H}$ together with a total-field/scattered-field (TF/SF) boundary, we introduced the incident beam in the material grid as shown in Fig. 3(b). This approach enables the introduction of the incident beam over a TF/SF boundary that parallels and is adjacent to the interface. (This permits the use of a much smaller computational domain than the one described in [9].)

The incident field is a pulsed Gaussian beam as described in [17] with beam waist $W_0 \approx 20\lambda$. Gaussian beams consist of a (angular) spectrum of plane waves where the angle of propagation of the spectral components is described by a Gaussian distribution (about the nominal incident angle θ_i). The larger the beam waist, the narrower the spectrum. The beam waist used here is slightly more than twice that used by Willis *et al.* [9].

During the simulations, snapshots were taken of the Poynting vector ($\mathbf{S} = \mathbf{E} \times \mathbf{H}$). In the 2-D TM^z grid, $\mathbf{S}$ has only two orthogonal components, S_x (perpendicular to the interface) and S_y (tangential to the interface). In the FDTD grid, S_x and S_y are represented by $S_x^q(m,n)$ and $S_y^q(m,n)$, where m is the spatial

step in the x-direction, n is the spatial step in the y-direction, and q is the time-step, such that

$$S_x^q(m,n) = -E_z^q(m,n)H_y^q(m,n), \quad (1)$$
$$S_y^q(m,n) = E_z^q(m,n)H_x^q(m,n). \quad (2)$$

However, because of the magnetic field's half-step offset in both space and time relative to the electric field, averaging is used to obtain the magnetic fields at the necessary temporal and spatial location. In Fig. 3, we defined the direction from the first medium to second medium (left to right) as the positive x direction. Thus, if S_x is positive, power flows toward the second medium while a negative value of S_x indicates power is flowing away from the second medium.

As indicated in Fig. 3(b), an observation line is placed in the grid [although drawn slightly to the left of the interface in Fig. 3(b), for the results to follow, the observation line is collocated with the interface]. At each time-step, we calculate the power flow "into" and "out of" the second medium, $P_{\text{in}}(q)$ and $P_{\text{out}}(q)$, respectively. These are given by

$$P_{\text{in}}(q) = \sum_n \frac{1+\text{sgn}\left(S_x^q(m_0,n)\right)}{2} S_x^q(m_0,n) \quad (3)$$
$$P_{\text{out}}(q) = \sum_n \frac{-1+\text{sgn}\left(S_x^q(m_0,n)\right)}{2} S_x^q(m_0,n) \quad (4)$$

where m_0 corresponds to the x index of the interface and $\text{sgn}()$ is the signum function. The range for n in the sum is from the bottom of the grid to the top, i.e., (3) and (4) perform a spatial integration over the interface at a fixed time-step. Integrating these values over time and taking their difference yields the "net" energy $\mathcal{P}(q)$ flow from the second medium back into the first medium

$$\mathcal{P}(q) = \sum_{k=0}^{q} P_{\text{out}}(k) - \sum_{k=0}^{q} P_{\text{in}}(k). \quad (5)$$

III. Results

The computational domain for the following simulations is 1000 (x) by 5000 (y) uniform cells, where the interface is at an x index 900, and the incident angle is 48°. Fig. 4 is a plot of $\mathcal{P}(q)$ when the second medium is gainy, lossless, or lossy. The solid curve, corresponding to the lossless case, is initially zero and, starting around 5000 time-steps, becomes increasingly negative as there is a net flow of energy to the right. After approximately 6800 time-steps, the curve stops decreasing and begins to climb. Eventually, the curve returns to zero, indicating there was no net flow of energy into or from the second medium: All the energy that passed to the right eventually passed back to the left. When the second medium is lossy, the curve again starts to drop at around 5000 time-steps, but the curve drops even further and never returns to zero, indicating that there has been a net flow of energy into the second medium. As with the other two media, the curve for the gainy medium also starts to drop around 5000 time-steps, indicating energy is flowing into the second medium. However, the drop is not as great as for the other two media, and eventually the curve moves into positive territory, indicating there has been a net flow of energy out of the second medium. *Significantly*, this curve does not obtain a positive value until after a significant amount of time has passed—the reflected field is not immediately greater than the incident field.

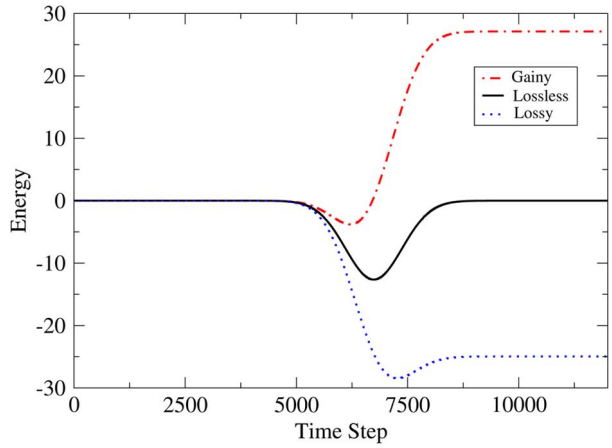

Fig. 4. Accumulated net energy flow from the second medium versus time for lossy, lossless, and gainy media.

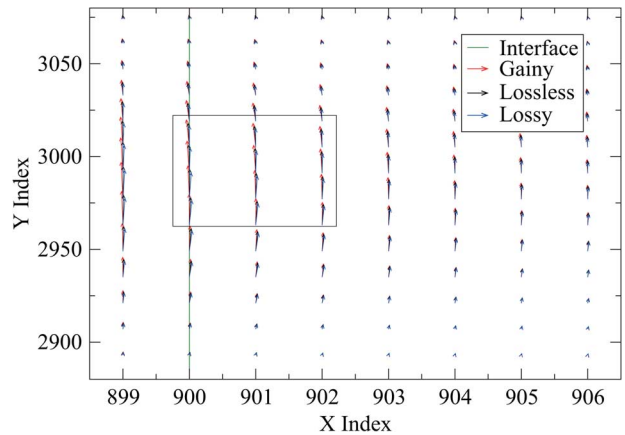

Fig. 5. Snapshot of Poynting vectors at time-step 6699 for lossless, lossy, and gainy media.

To provide a qualitative understanding of the behavior of the fields, we examined snapshots of the Poynting vectors near the interface. One such snapshot (taken at time step 6699) is shown in Fig. 5. A subset of the grid is shown where samples are taken every 14 points along the y-axis (ranging from 2880 to 3080) and at every point along the x-axis (ranging from 899 to 906). Here, the conductivities for the gainy and lossy materials are -1500 and 1500 S/m, respectively (the magnitude of the conductivity was increased by a factor of three in order to make its affect on the Poynting vectors easier to see at a glance). One observes that the amount of "tilt" of the Poynting vectors in the direction of the first medium is greatest for the gainy material and least for the lossy material, indicating that the gainy material will have the most energy flow back toward the first medium while the lossy material will have the least flow in that direction (which is consistent with Fig. 4). Fig. 6 provides an expanded view of the rectangular box in Fig. 5 and further expands the view of one representative vector cluster.

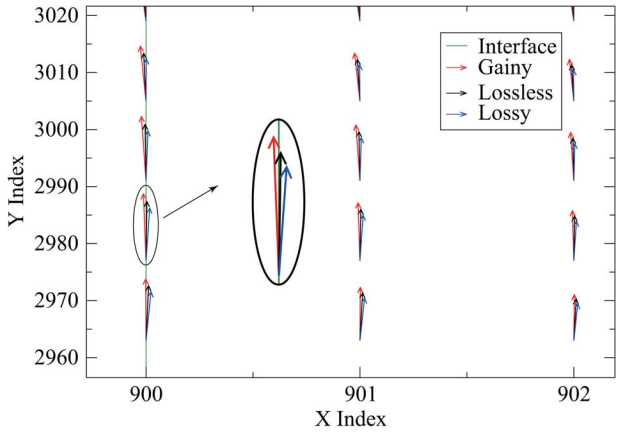

Fig. 6. Expanded view of the rectangular region shown in Fig. 5.

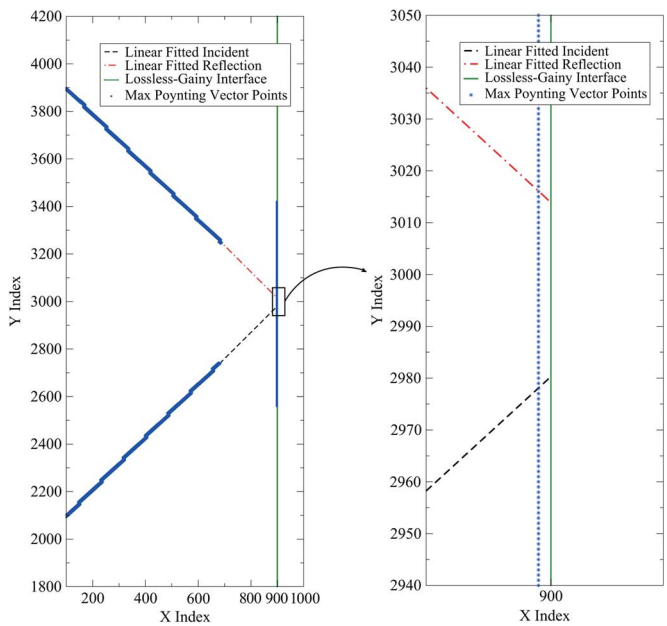

Fig. 7. Depiction of the Goos–Hänchen shift between the incident and reflected fields as obtained by breadcrumbing the maximum value of the Poynting vector when it is propagating toward or away from the interface.

To demonstrate the Goos–Hänchen shift inherent in TIR, we recorded the path of the maximum energy associated with the incident beam and the reflected beam as shown in Fig. 7. This path is obtained by finding the location of the maximum value of the Poynting vectors at every time-step and placing a "breadcrumb" at that location. This path starts at the TF/SF boundary and propagates toward the interface. As the Gaussian beam nears the lossless–gainy interface, the location of the peak energy transitions to the interface and propagates tangentially along it. As time progresses, the peak energy moves from the interface and follows the path of the reflected beam as shown in Fig. 7. We use breadcrumbs to fit straight lines to the path of the incident and reflected fields. As shown in Fig. 7, there is a small displacement between where the incident and reflect paths intersect the interface. This displacement is the Goos–Hänchen shift. (In this particular simulation, the displacement is approximately 33.6δ, where δ is the spatial step-size of the uniform grid.) To account for this shift, fields must "travel" within the second medium. When the second medium is gainy, this "travel" provides the opportunity for the fields to grow, hence experience gain, resulting in a reflected field that carries more energy than the incident field. Note that the Goos–Hänchen shift is: 1) a well-established fact of TIR; 2) a discontinuous phenomenon in that it is not present for incident angles below the critical angle and is maximum just beyond the critical angle; and 3) a decreasing function for increasing incident angles. Similarly, for a gainy material, the magnitude of the reflection coefficient is discontinuous at the critical angle (with a peak just beyond the critical angle) and decreases for increasing incident angles beyond the critical angle.

IV. Conclusion

FDTD simulations of TIR from a gainy material yield a reflection coefficient with a magnitude greater than unity. We have shown that, from the standpoint of energy flow, this gain should be rather unsurprising. The "excess" reflected energy only appears *after* the fields have had a chance to interact with the second medium. The behavior of the magnitude of the reflected field is also completely consistent with the Goos–Hänchen shift.

References

[1] K. Kao and G. Hockham, "Dielectric-fibre surface waveguides for optical frequencies," *Proc. IEE*, vol. 113, no. 7, pp. 1151–1158, 1966.
[2] C. Tsai, B. Kim, and F. El-Akkari, "Optical channel waveguide switch and coupler using total internal reflection," *IEEE J. Quantum Electron.*, vol. QE-14, no. 7, pp. 513–517, Jul. 1978.
[3] D. Axelrod, "Cell-substrate contacts illuminated by total internal reflection fluorescence," *J. Cell Bio.*, vol. 89, no. 1, pp. 141–145, 1981.
[4] J. Y. Han, "Low-cost multi-touch sensing through frustrated total internal reflection," in *Proc. 18th ACM UIST*, New York, NY, USA, 2005, pp. 115–118.
[5] H. J. Shaw, M. J. Digonnet, and R. A. Lacy, "Fiber optic amplifier," US Patent 4,553,238, Nov. 12, 1985.
[6] P. S. Epstein, "Reflection of waves in an inhomogeneous absorbing medium," *Proc. Nat. Acad. Sci.*, vol. 16, no. 10, pp. 627–637, 1930.
[7] J. R. F. Cybulski and C. K. Carniglia, "Internal reflection from an exponential amplifying region," *J. Opt. Soc. Amer.*, vol. 67, no. 12, pp. 1620–1627, Dec. 1977.
[8] J. Fan, A. Dogariu, and L. J. Wang, "Amplified total internal reflection," *Opt. Express*, vol. 11, no. 4, pp. 299–308, Feb. 2003.
[9] K. J. Willis, J. B. Schneider, and S. C. Hagness, "Amplified total internal reflection: Theory, analysis, demonstration of existence via FDTD," *Opt. Express*, vol. 16, no. 3, pp. 1903–1914, Feb. 2008.
[10] A. Siegman, "Fresnel reflection, Lenserf reflection and evanescent gain," *Opt. Photon. News*, vol. 21, no. 1, pp. 38–45, Jan. 2010.
[11] C. Koester, "Laser action by enhanced total internal reflection," *IEEE J. Quantum Electron.*, vol. 2, no. 9, pp. 580–584, 1966.
[12] G. N. Romanov and S. S. Shakhidzhanov, "Amplification of electromagnetic field in total internal reflection from a region of inverted population," *Sov. J. Exp. Theor. Phys. Lett.*, vol. 16, p. 210, Sep. 1972.
[13] S. Lebedev, V. Volkov, and B. Y. Kogan, "Value of the gain for light internally reflected from a medium with inverted population," *Opt. Spectrosc.*, vol. 35, pp. 565–566, 1973.
[14] P. R. Callary and C. K. Carniglia, "Internal reflection from an amplifying layer," *J. Opt. Soc. Amer.*, vol. 66, no. 8, pp. 775–779, Aug. 1976.
[15] F. Goos and H. Hänchen, "Ein neuer und fundamentaler versuch zur totalreflexion," *Ann. Phys.*, vol. 436, no. 7–8, pp. 333–346, 1947.
[16] J. A. Roden and S. D. Gedney, "Convolution PML (CPML): An efficient FDTD implementation of the CFS-PML for arbitrary media," *Microw. Opt. Technol. Lett.*, vol. 27, no. 5, pp. 334–339, 2000.
[17] Z. Wang, Z. Zhang, Z. Xu, and Q. Lin, "Space-time profiles of an ultrashort pulsed Gaussian beam," *IEEE J. Quantum Electron.*, vol. 33, no. 4, pp. 566–573, Apr. 1997.