

On Fresnel Reflection and Evanescent Gain

Аннотация

We point and rectify several dubious and incorrect claims, listed in one of the articles published by OPN [1].

1 Introduction

After reading of paper [1], we thought it could be a New Year's prank aimed at testing readers's intelligence. In this manuscript we would like to reveal, as the author of [1] puts it, "dubious, whimsical or even incorrect" statements of [1] to prove that the OSA members and other readers of the OPN magazine maintain high level of alertness. Our text has significant technical content, which is however necessary to properly convey our rebuttal of [1].

2 Maxwell and wave equations

We begin by recalling the Maxwell's equations in a medium without free charges, with zero conductivity σ , and time-independent permittivities ϵ , μ (see [2], for example):

$$-[\nabla \times \mathbf{E}(\mathbf{r}, t)] = \mu \frac{\partial}{\partial t} \mathbf{H}(\mathbf{r}, t), \quad (1)$$

$$[\nabla \times \mathbf{H}(\mathbf{r}, t)] = \epsilon \frac{\partial}{\partial t} \mathbf{E}(\mathbf{r}, t) \quad (2)$$

$$\nabla \cdot \epsilon \mathbf{E}(\mathbf{r}, t) = 0, \quad \nabla \cdot \mu \mathbf{H}(\mathbf{r}, t) = 0. \quad (3)$$

If ϵ and μ in addition are constant in space, the last two equations simplify to

$$\nabla \cdot \mathbf{E}(\mathbf{r}, t) = 0, \quad \nabla \cdot \mathbf{H}(\mathbf{r}, t) = 0, \quad (4)$$

and with the help of (1), (2) and (4) we can obtain the wave equations for the fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{H}(\mathbf{r}, t)$ as shown below.

The time derivative of both parts of (2), taking into account (1), gives

$$-[\nabla \times [\nabla \times \mathbf{E}(\mathbf{r}, t)]] = \mu \epsilon \frac{\partial^2}{\partial t^2} \mathbf{E}(\mathbf{r}, t), \quad (5)$$

which, with the help of (4), is reduced to the well known wave equation for the electric field $\mathbf{E}(\mathbf{r}, t)$:

$$\Delta \mathbf{E}(\mathbf{r}, t) = -\mu \epsilon \frac{\partial^2}{\partial t^2} \mathbf{E}(\mathbf{r}, t). \quad (6)$$

Similarly, time derivative of both parts of (1), and using (2), gives the wave equation for the magnetic field $\mathbf{H}(\mathbf{r}, t)$:

$$\Delta \mathbf{H}(\mathbf{r}, t) = -\mu \epsilon \frac{\partial^2}{\partial t^2} \mathbf{H}(\mathbf{r}, t). \quad (7)$$

Both equations have solutions in the form of a plain wave

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E} \exp(i\mathbf{k}\mathbf{r} - i\omega t), \quad \mathbf{H}(\mathbf{r}, t) = \mathbf{H} \exp(i\mathbf{k}\mathbf{r} - i\omega t). \quad (8)$$

By placing these solutions into the wave equations we find that $k^2 = \epsilon\mu\omega^2 = \omega^2/c^2$, where $c = 1/\sqrt{\epsilon\mu}$ is the speed of light in the medium.

The two wave equations are independent, and it may appear at first that we can consider both fields also independently and represent the full electromagnetic field by the wave function

$$\boldsymbol{\psi}(\mathbf{r}, t) = (\mathbf{E} + \mathbf{H}) \exp(i\mathbf{k}\mathbf{r} - i\omega t), \quad (9)$$

with a single restriction that both \mathbf{E} and \mathbf{H} are perpendicular to \mathbf{k} . However, the wave equations are derived from the Maxwell's equations, and the outcome of the Maxwell's equations after substitution of the first (8) into (1) is

$$\mathbf{H} = \frac{1}{\mu\omega}[\mathbf{k} \times \mathbf{E}], \quad (10)$$

or similarly, after substitution of the second (8) into (2) is

$$\mathbf{E} = -\frac{1}{\epsilon\omega}[\mathbf{k} \times \mathbf{H}]. \quad (11)$$

Therefore, \mathbf{E} and \mathbf{H} are not independent. They are orthogonal to each other and if $|\mathbf{E}| = 1$, the length of \mathbf{H} is $|\mathbf{H}| = \sqrt{\epsilon/\mu} = 1/Z$, where $Z = \sqrt{\mu/\epsilon}$ is called the medium impedance.

3 Wave reflection and refraction at an interface

If space is not completely homogeneous but consists of two parts with different ϵ and μ , we cannot replace (3) with (4), because the permittivities now depend on the coordinates. However any part of space is homogeneous, and each part has its own wave equation with its own plain wave solution. The transmission through an interface is similar to travelling through a border between two foreign countries where certain rules are imposed. In case of two media, the rules are imposed by Maxwell's equations. According to these rules, the transmission is allowed only if (according to (2) and (1)) components $\mathbf{E}_{\parallel}(\mathbf{r}, t)$, $\mathbf{H}_{\parallel}(\mathbf{r}, t)$ of the fields $\mathbf{E}(\mathbf{r}, t)$, $\mathbf{H}(\mathbf{r}, t)$ parallel to the interface are the same, and (according to (3)) the products $\epsilon(\mathbf{n} \cdot \mathbf{E}(\mathbf{r}, t))$, $\mu(\mathbf{n} \cdot \mathbf{H}(\mathbf{r}, t))$, where \mathbf{n} is a unit vector \mathbf{n} normal to the interface, are the same on both sides of the border/interface. Because of these restrictions only a fraction (refracted, or transmitted) of the incident wave is permitted to go through the border, and the remaining part (reflected) is ordered to go back. So, before starting our journey we should calculate how much of the incident wave (9) will be transmitted and how much will be reflected at an interface between space 1 at $z < 0$ and space 2 at $z > 0$, which have different electromagnetic constants $\epsilon_{1,2}$ and $\mu_{1,2}$. We can expect that the total wave function in presence of interface looks

$$\boldsymbol{\psi}(\mathbf{r}, t) = \Theta(z < 0) \left(\exp(i\mathbf{k}_1\mathbf{r} - i\omega t)\boldsymbol{\psi}_1 + \exp(i\mathbf{k}_r\mathbf{r} - i\omega t)\boldsymbol{\psi}_r \right) + \Theta(z > 0) \exp(i\mathbf{k}_2\mathbf{r} - i\omega t)\boldsymbol{\psi}_2\tau, \quad (12)$$

where the term $\exp(i\mathbf{k}_1\mathbf{r} - i\omega t)\boldsymbol{\psi}_1$ with the wave vector \mathbf{k}_1 describes the plain wave incident on the interface from medium 1, factors $\boldsymbol{\psi}_i = \mathbf{E}_i + \mathbf{H}_i$ ($i = 1, r, 2$) denote sum of electric and magnetic polarization vectors, $\mathbf{k}_{r,2}$ are wave vectors of the reflected and transmitted waves, ρ , τ are reflection and transmission amplitudes respectively, and $\Theta(z)$ is the step function, which is equal to unity when inequality in its argument is satisfied, and to zero otherwise.

3.1 Wave vectors of reflected and refracted waves

The wave vectors $\mathbf{k}_{r,2}$ of the reflected and refracted waves are completely determined by the wave vector \mathbf{k}_1 of the wave incident on the interface from the medium 1. And at this point we arrive at **the first dubious claim** of [1]:

1. (page 40, right column, line 5 from the bottom) Satisfying the boundary conditions for the E and H fields at the interface then provides a recipe for finding \mathbf{k}_2 as a function of the incident \mathbf{k}_1 .

Contrary to the above statement, the recipe for finding \mathbf{k}_2 as a function of the incident \mathbf{k}_1 is entirely independent of the boundary conditions. It is determined uniquely by the constants ϵ_i , μ_i and by the fact that the frequency ω and part \mathbf{k}_\parallel along the interface of the wave vectors must be identical for all the waves. In the following we assume that medium 1 is lossless, i.e. $\epsilon_1\mu_1$ is real. In this case all the components of the wave vector \mathbf{k}_1 are also real.

Frequency of all the waves is the same because the reflection and refraction is an elastic scattering process. Vector \mathbf{k}_\parallel is constant because the space along the interface is homogeneous and there are no points where \mathbf{k}_\parallel could change. Thus, the length of the refracted wave vector is $|\mathbf{k}_2| = \omega\sqrt{\epsilon_2\mu_2}$, which can also be written as

$$|\mathbf{k}_2| = \omega\sqrt{\epsilon_1\mu_1}\sqrt{\frac{\epsilon_2\mu_2}{\epsilon_1\mu_1}} = k_1n = k_1\sqrt{\epsilon}, \quad (13)$$

where n is the relative refraction index,

$$n = \sqrt{\frac{\epsilon_2\mu_2}{\epsilon_1\mu_1}} \quad (14)$$

and where we introduced $\epsilon = n^2$, as used for light refraction from vacuum into a dielectric.

Because \mathbf{k}_\parallel is the same for all the waves, and the medium 1 is isotropic, the reflection is specular. Therefore, the normal component of the reflected wave vector is $k_{r\perp} = -k_{1\perp} = -\sqrt{k_1^2 - \mathbf{k}_\parallel^2}$. The normal component of the refracted wave vector is

$$k_{2\perp} = \sqrt{\epsilon k_1^2 - \mathbf{k}_\parallel^2}. \quad (15)$$

Thus, one can see that we do not use the boundary conditions for \mathbf{E} and \mathbf{H} fields to obtain the above results.

3.1.1 Medium 2 is lossless

From the latter expression it follows that for lossless media (ϵ is real) when $0 < \epsilon < 1$, the incident wave with \mathbf{k}_\parallel in the interval $nk_1 \leq |\mathbf{k}_\parallel| \leq k_1$ is totally reflected from the interface, because

$$k_{2\perp} = iK_{2\perp}'' \equiv i\sqrt{k_\parallel^2 - \epsilon k_1^2}, \quad (16)$$

the factor $\exp(ik_{2\perp}z) = \exp(-K_{2\perp}''z)$ of the wave $\exp(i\mathbf{k}_2\mathbf{r})$ exponentially decays, and the refracted wave becomes an evanescent one. Therefore, the energy does not flow inside the medium 2, and because of the energy conservation it must be totally reflected back into medium 1. Let us note, however, that the transition of the propagating wave to evanescent one is continuous, because at the transition point $|\mathbf{k}_\parallel| = nk_1$ we have $k_{2\perp} = K_{2\perp}'' = 0$. The latter is a correction of **the second dubious claim** in [1], which using the above notations can be written as

2.(page 41 left column line 7) the behavior of $k_{2\perp}$ changes “discontinuously” to give $k_{2\perp} = 0$ while the vertical component of exponential decay vector becomes imaginary.

3.1.2 Medium 2 is lossy or gainy

If the medium 2 is lossy or gainy, the constant ϵ is a complex quantity $\epsilon = \epsilon' \pm i\epsilon''$ with positive ϵ' and ϵ'' . In this case, far outside the total internal reflection (TIR) region ($|\mathbf{k}_{\parallel}|^2 \ll \epsilon'k_1^2$) for small ϵ'' ($\epsilon''k_1^2 \ll \epsilon'k_1^2 - |\mathbf{k}_{\parallel}|^2$) we have

$$k_{2\perp} = \sqrt{\epsilon'k_1^2 - |\mathbf{k}_{\parallel}|^2 \pm i\epsilon''k_1^2} = k'_{2\perp} \pm ik''_{2\perp}, \quad (17)$$

where

$$k'_{2\perp} \approx \sqrt{\epsilon'k_1^2 - |\mathbf{k}_{\parallel}|^2}, \quad k''_{2\perp} \approx \epsilon'' \frac{k_1^2}{2k'_{2\perp}}. \quad (18)$$

The sign of the square root in (17) is positive, because $k'_{2\perp}$ must be positive to describe propagation of the refracted wave away from the interface. At the same time we see that for lossy medium the imaginary part should have positive sign to get exponential decay of the refracted wave; and for gainy media it should have negative sign to get exponential growth of the refracted wave. Here we arrive at another **erroneous claim** (we remind that the imaginary part n'' of the refractive index $n = n' + in'' = \sqrt{\epsilon' \pm i\epsilon''}$ has the same sign as the sign before ϵ''):

3. (page 41, right column, line 7 from the bottom) imaginary part corresponds to absorption loss if n'' is negative, and to gain if n'' is positive.

Now let's look at TIR. Here, $k'_{2\perp}$ in (17) transforms into $iK''_{2\perp}$, where $K''_{2\perp} \approx \sqrt{|\mathbf{k}_{\parallel}|^2 - \epsilon'k_1^2}$, and $k''_{2\perp}$ in (18) transforms to

$$k''_{2\perp} \rightarrow -iK'_{2\perp} = \epsilon'' \frac{k_1^2}{2iK''_{2\perp}}. \quad (19)$$

Therefore, in TIR we have

$$k_{2\perp} = \sqrt{-(|\mathbf{k}_{\parallel}|^2 - \epsilon'k_1^2) \pm i\epsilon''k_1^2} = \pm K'_{2\perp} + iK''_{2\perp}, \quad (20)$$

where

$$K'_{2\perp} = \epsilon'' \frac{k_1^2}{2K''_{2\perp}}, \quad K''_{2\perp} \approx \sqrt{|\mathbf{k}_{\parallel}|^2 - \epsilon'k_1^2}. \quad (21)$$

(In the paper [1] $K'_{2\perp}$ is denoted k_{2x} , and $K''_{2\perp}$ is denoted $-g_{2x}$.) Again we see, that the sign before the square root is defined in the way to obtain the exponential decay of the refracted wave away from the interface for both lossy and gainy media cases. However the real part has opposite signs. The positive value of $K'_{2\perp}$ for lossy medium means that the reflection coefficient in TIR is less than unity, because part of the energy flux, proportional to $K'_{2\perp}$, enters the medium 2 and decays there. The negative value of $K'_{2\perp}$ for gainy medium means that the reflection coefficient in TIR is larger than unity, because part of the energy flux, proportional to $K'_{2\perp}$, exits the medium 2 and adds to the totally reflected wave. And here we arrive at two more **erroneous claims**, which in the above notations can be written as

4. (page 42, left column, line 6) k_{2x} taking on the same positive sign for equal values of either loss or gain,
5. (page 42, left column, line 7) while g_{2x} takes on positive and negative signs for the gainy and lossy cases respectively.

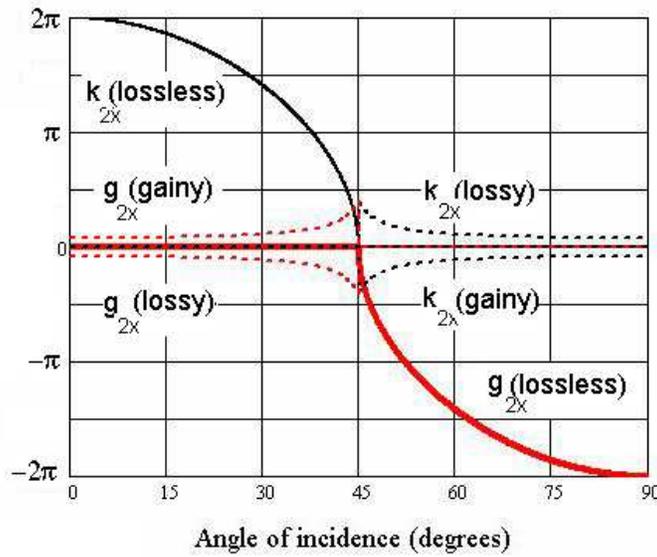


Рис. 1: This is how the Figure 2 of [1] should look like. The value of g_{2x} (in our notation it is $-K''_{2\perp}$) for lossy and gainy media in TIR, as well as k_{2x} (in our notation it is $K'_{2\perp}$) outside of it, are very close to the lossless case. The figure is calculated for wavelength $\lambda = 1$, $\epsilon' = 0.5$ (to start TIR at $\theta = 45^\circ$) and $\epsilon'' = 0.04$.

It is physically impossible for the wave function of the primary photon to exponentially increase in gainy medium. If it were so, it would mean that even in the case of infinitesimal ϵ'' we would have enormous increase in energy inside the gainy medium, and the smaller is the penetration of the incident photon, the larger is the increase of the penetrated radiation. The correct result shows that the less is the penetration of the primary wave into the medium 2, the less is the gain. The sign of $K'_{2\perp}$ is uniquely determined by the sign before ϵ'' and by the sign of $K''_{2\perp}$. Since the sign of $K''_{2\perp}$ for both types of media is the same, the sign of $K'_{2\perp}$ is different for lossy and gainy media. It follows that the Figure 2 of [1] should be replaced by our Fig.1, and all of the Figure 4 of [1] must be withdrawn.

3.2 Reflection and refraction amplitudes

The procedure for calculating the reflection amplitudes in general case is well explained in [3], so here we only briefly recall it. The polarization \mathbf{E}_1 of the incident wave can be arbitrary (except it must be perpendicular to the wave-vector \mathbf{k}_1), and it always can be decomposed as $\mathbf{E} = \mathbf{E}_{1s} + \mathbf{E}_{1p}$, where \mathbf{E}_{1s} is the component parallel to the interface and perpendicular to the plane of incidence (plane of vectors \mathbf{k}_1 and the normal \mathbf{n} to the interface), and where \mathbf{E}_{1p} lies in the plane of incidence. Reflection of each component are different and can be found independently.

Lets find the reflection of the wave \mathbf{E}_{1s} . It is usually called s-wave or TE-wave. The field \mathbf{E}_{1s} is accompanied by field \mathbf{H}_{1p} which lies in the incidence plane. The total wave function of the TE- wave, according to (12) can be represented as $\exp(i\mathbf{k}_{\parallel}\mathbf{r}_{\parallel} - i\omega t)\psi(z)$, where

$$\psi(z) = \Theta(z < 0) [\psi_{1s} \exp(ik_{1\perp}z) + \psi_{rs}\rho_s \exp(-ik_{1\perp}z)] + \Theta(z > 0) \psi_{2s}\tau_s \exp(ik_{2\perp}z), \quad (22)$$

where for $i = 1, r, 2$

$$\psi_{is} = \mathbf{E}_{1s} + \mathbf{H}_{ip}, \quad \mathbf{H}_{ip} = \frac{1}{\mu_i\omega} [\mathbf{k}_i \times \mathbf{E}_{1s}], \quad \mu_r = \mu_1, \quad (23)$$

and the corresponding wave vectors are

$$\mathbf{k}_1 = \mathbf{k}_{\parallel} + \mathbf{n}k_{1\perp}, \quad \mathbf{k}_r = \mathbf{k}_{\parallel} - \mathbf{n}k_{1\perp}, \quad \mathbf{k}_2 = \mathbf{k}_{\parallel} + \mathbf{n}k_{2\perp}. \quad (24)$$

Maxwell's equations require continuity at the interface of the electric field \mathbf{E}_s , which leads to the equation $1 + \rho_s = \tau_s$. The requirement for the parallel to the interface component $\mathbf{H}_{\parallel p}$ of the magnetic field to be continuous leads to the equation $(1 - \rho_s)k_{\perp 1}/\mu_1 = \tau_s k_{\perp 2}/\mu_2$. The third requirement for the continuity of the quantity $\mu(\mathbf{n} \cdot \mathbf{H}_p)$ leads to the same equation $1 + \rho_s = \tau_s$ as the one obtained from the continuity of \mathbf{E}_s . Therefore we have two independent equations from which we obtain the well known Fresnel formulas

$$\rho_s = \frac{\mu_2 k_{1\perp} - \mu_1 k_{2\perp}}{\mu_2 k_{1\perp} + \mu_1 k_{2\perp}}, \quad \tau_s = \frac{2\mu_2 k_{1\perp}}{\mu_2 k_{1\perp} + \mu_1 k_{2\perp}}. \quad (25)$$

Similar considerations of the TH-wave with \mathbf{E}_{1p} polarization gives two other expressions

$$\rho_p = \frac{\epsilon_2 k_{1\perp} - \epsilon_1 k_{2\perp}}{\epsilon_2 k_{1\perp} + \epsilon_1 k_{2\perp}}, \quad \tau_p = \frac{2\epsilon_2 k_{1\perp}}{\epsilon_2 k_{1\perp} + \epsilon_1 k_{2\perp}}. \quad (26)$$

As in [1] we limit ourselves only to TE-case and assume that $\mu_2 = \mu_1$, so the formulas (25) are reduced to

$$\rho_s = \frac{k_{1\perp} - k_{2\perp}}{k_{1\perp} + k_{2\perp}}, \quad \tau_s = \frac{2k_{1\perp}}{k_{1\perp} + k_{2\perp}}. \quad (27)$$

Now we can use these expressions to draw the polar graph that corresponds to the Figure 3 in [1], but first we will mention **three more erroneous claims**, written in the above notations

6. (page 42, left column, line 22) As the incidence angle passes through what would have been the critical angle for the lossless case the contours for the lossy and gainy amplitudes... progress in opposite directions around the unit circle.

7. (page 42, middle of left column) The primary result is that $\rho_s(\theta)$ becomes less than unity for both lossy and gainy cases at all angles of incidence, including angles within TIR.

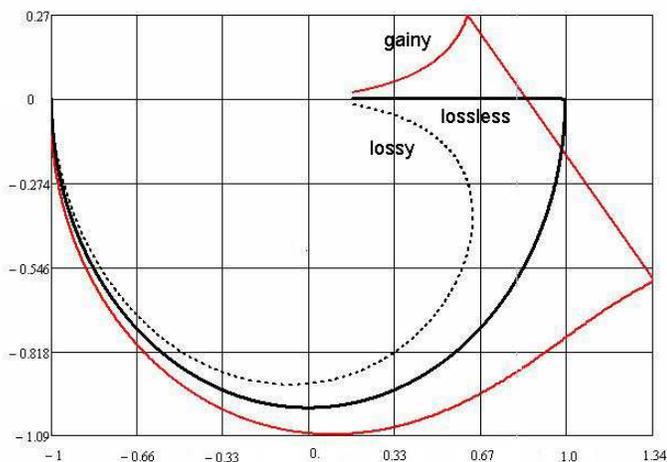


Рис. 2: This is how the Figure 3 of [1] should look like. Here, the vertical and horizontal axes correspond to the imaginary and real parts of the amplitude ρ_s of Eq.(27) respectively. We see that the imaginary part at TIR is always negative for both medias, contrary to results presented in Fig. 3 of [1]. The figure is calculated for $\epsilon' = 0.5$ and $\epsilon'' = 0.04$.

8. (page 42, left column, line 15 from bottom) The magnitude of the reflection coefficient always remains less than unity for any angle of incidence and for any value of loss or gain.

Fig. 2, which should replace the Figure 3 of [1], shows that within TIR the contours of the reflection amplitudes for all the cases progress in the same direction around the unit circle. It shows that in TIR the reflection coefficient from a gainy medium is larger than unity and increases with gain. It seems that the grows of the reflection coefficient can be explained by the photon emission, induced by the evanescent field toward the interface. The induced photon cannot propagate inside the gainy medium within TIR, for the same reasons as the primary photon cannot propagate. Therefore the increase in the reflected flux is due to the sub-barrier induction of the photon, which tunnels from the gainy medium into vacuum and coherently adds to the reflected primary photon. The larger is the gain, the larger is the probability of such process. The difference $|\rho_s|^2 - 1$ is shown in Fig. 3. We see that the magnitude of the reflection coefficient in TIR is always larger than unity contrary to the claim in the publication [1].

3.3 On definition of the energy flux

We also want to note that the widely spread belief that the energy flux is given by the Pointing vector $\mathbf{J} = [\mathbf{E} \times \mathbf{H}]$ is not correct. The energy flux is given by

$$\mathbf{J} = c \frac{\mathbf{k}}{k} \frac{\epsilon E^2 + \mu H^2}{8\pi}, \quad (28)$$

i.e. it is equal to energy density times the light speed, and it has a direction along the wave vector \mathbf{k} . For a plain wave this definition coincides with the Pointing vector. However the latter can be defined for wider variety of vectors \mathbf{E} and \mathbf{H} , even for stationary fields, where it has no relation to the energy flux. The Pointing vector can also be written for an evanescent wave, but in this case it does not correspond to the energy flux.

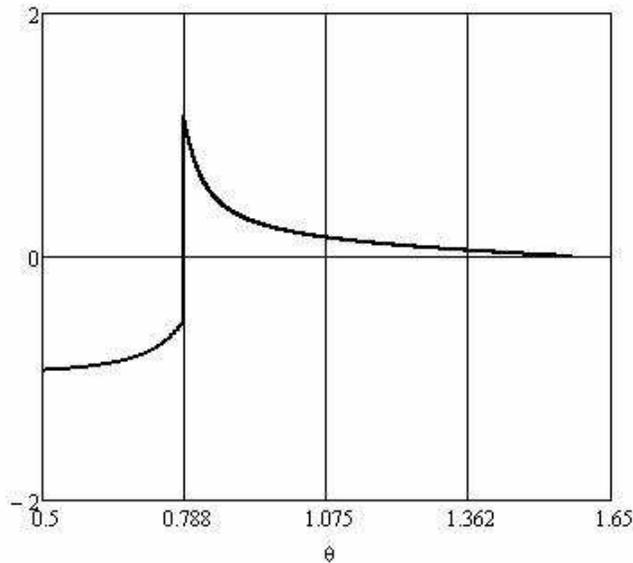


Рис. 3: The difference of $|\rho_s(\theta)|^2 - 1$ near and within TIR, calculated with the same parameters as Fig. 2.

3.4 The Goos-Hänchen (G-H) shift

Here again we meet another **dubious claim** in [1]:

9. (page 42, left column, line 12 from bottom) Readers interested in the Goos-Hänchen shift will note that this shift now evidently takes on equal but opposite values for the lossy and gainy cases.

The above statement gives impression, that there will be no G-H shift for lossless media. However it is not true. The Goos-Hänchen (G-H) shift, though not uniquely defined, has a value, which in first approximation does not depend on the sign before ϵ'' . One of the definitions is given in [4, 5], where the incident wave is represented by a pencil-ray, and the G-H shift for lossless media is defined as the distance along the interface between the centers of the footprints of the incident and reflected pencil-rays. This distance is then calculated to be

$$d_{\parallel} = 2d\varphi'/dk_{\parallel} = \frac{2k_{\parallel}}{k_{\perp}\sqrt{k_{\parallel}^2 - \epsilon k^2}}. \quad (29)$$

Thus, the G-H shift in this definition is positive, its sign does not depend on the sign of the imaginary part of the refractive index, and therefore is the same for lossless, lossy and gainy media.

4 Conclusion

We listed here most of the erroneous claims of [1], to show that we are worth the free OSA membership. We avoided the discussion of the Lensef formulas, because these formulas give impression of a kids game with the signs of the square roots, without any attention to the physical meaning. We also did not discuss the experimental details, however we are open to discussion, if anyone is interested. Overall, we admired reading such a vibrant paper, and we are happy to realize that we are still capable of passing a college physics exam, despite the long time since the graduation.

Список литературы

[*] e-mail: v.ignatovi@gmail.com

- [1] A.Siegman, Fresnel Reflection, and Evanescent Gain, OPN, V. 21, n. 1, 38-45 (2010).
- [2] Javier E. Vitela, Electromagnetic waves in dissipative media revisited, Am. J. Phys. 72, 393-403 (2004).
- [3] L. D. Landau, E. M. Lifshitz and L. P. Pitaevskii, Electrodynamics of Continuous Media. Second Edition: Volume 8 (Course of Theoretical Physics) Elsevier Butterworth-Heinemann, 2004. Ch.X, § 86.
- [4] H. Hora, Optik 17, 409 (1960).
- [5] V. Ignatovich, Phys. Lett. A 322, 36 (2004).