Fresnel reflection from a cavity with net roundtrip gain

Tobias S. Mansuripur[1,*] and Masud Mansuripur[2]

[1] Department of Physics, Harvard University, Cambridge, MA 02139
[2] College of Optical Sciences, The University of Arizona, Tucson, AZ 85721

A planewave incident on an active etalon with net roundtrip gain may be expected to diverge in field amplitude, yet Maxwell's equations admit only a convergent solution. By examining a Gaussian beam obliquely incident on such a cavity, we find that the "side-tail" of the beam leaks into the cavity and gives rise to a field that interferes with the main portion of the beam, which is ultimately responsible for the convergence of the field. This mechanism offers perspective for many phenomena, and we specifically discuss the implications for amplified total internal reflection.

The Fresnel coefficients govern the reflection and transmission of light for the simplest possible scenarios: at planar interfaces between homogeneous media. Despite this simplicity, some interesting solutions have been discovered only recently, such as 1) the amplification of evanescent waves in a passive, negative-index slab [1–3], and controversy regarding the proper choice of the wavevector in active media has persisted in relation to the possibility of 2) negative refraction in nonmagnetic media [4–8] as well as 3) single-surface amplified total internal reflection (TIR) [9–14]. It turns out that all three of these cases share a common thread: the presence of a cavity whose roundtrip gain exceeds the loss. In this Letter, we explore more generally the response of such a cavity to an incident beam of light.

To begin, we establish conventions that allow us to more clearly discuss the direction of energy flow, which is central to our overall argument. For the single-surface problem, shown in Fig. 1(a), the incident wavevector in medium one is $\boldsymbol{k}_1^R = k_x\hat{\boldsymbol{x}} + k_{1z}^R\hat{\boldsymbol{z}}$, and the reflected wavevector is $\boldsymbol{k}_1^L = k_x\hat{\boldsymbol{x}} + k_{1z}^L\hat{\boldsymbol{z}}$, where $k_{1z}^L = -k_{1z}^R$. The component k_x (which we assume for simplicity to be real-valued), once determined by the incident wave, must become the x-component of every wavevector in the system in order to satisfy Maxwell's boundary conditions. For the transmitted wavevector $\boldsymbol{k}_2$, the dispersion relation offers two choices for k_{2z},

$$k_{2z} = \pm\sqrt{(\omega/c)^2\mu_2\epsilon_2 - k_x^2}, \tag{1}$$

where ω is the angular frequency of the planewave and c is the speed of light in vacuum. We denote by k_{2z}^R (k_{2z}^L) the choice which carries energy to the right (left), namely, that for which the time-averaged z-component of the Poynting vector is positive (negative). (In cases where both choices for k_{2z} result in no energy flow in the z-direction, such as for evanescent waves in a transparent medium, our prescription is to add a small amount of loss to the slab which will unambiguously distinguish k_{2z}^R and k_{2z}^L, then take the limit as the loss goes to zero. See the supplemental information for details.) Let us

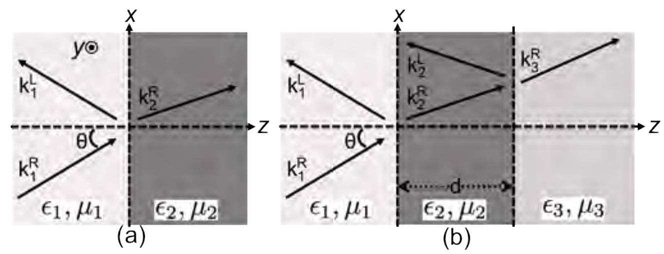

FIG. 1. Geometry of the (a) single-surface and (b) cavity problems. All media are infinite in the x and y-directions. The arrows denote the wavevectors of the planewaves present in each layer. The material constants are the relative permittivities and permeabilities.

postulate that the proper choice for k_{2z} in the single-surface problem is always k_{2z}^R (i.e., that the transmitted energy flows away from the interface), irrespective of the material parameters or the nature of the incident wave. (In fact, k_{2z}^R is universally agreed to be the correct choice in all cases except possibly that of amplified TIR; it is due to this one controversy that we refer to this choice as a postulate for now.) We require this postulate in order to unambiguously define the single-surface Fresnel reflection and transmission coefficients

$$r_{\ell m} = \frac{\tilde{k}_{\ell z}^R - \tilde{k}_{mz}^R}{\tilde{k}_{\ell z}^R + \tilde{k}_{mz}^R},\ t_{\ell m} = \frac{2\tilde{k}_{\ell z}^R}{\tilde{k}_{\ell z}^R + \tilde{k}_{mz}^R} \tag{2}$$

where we have generalized the result for incidence medium ℓ and transmission medium m. For s-polarization we have defined $\tilde{k}_{nz} \equiv k_{nz}/\mu_n$, and the two coefficients yield the reflected and transmitted amplitudes of the component E_y, while for p-polarization $\tilde{k}_{nz} \equiv k_{nz}/\epsilon_n$ and the coefficients are associated with the component E_x.

Having established these conventions, we now consider the case of light incident on a cavity, shown in Fig. 1(b). The total E-field resulting from an s-polarized incident

* mansuripur@physics.harvard.edu

wave in medium one with amplitude E_1^R is given by

$$E_y(x,z) = \begin{cases} E_1^R \exp(ik_x x + ik_{1z}^R z) \\ \quad +E_1^L \exp(ik_x x + ik_{1z}^L z) & : z \leq 0 \\ E_2^R \exp(ik_x x + ik_{2z}^R z) \\ \quad +E_2^L \exp(ik_x x + ik_{2z}^L z) & : 0 \leq z \leq d \\ E_3^R \exp[ik_x x + ik_{3z}^R (z-d)] & : z \geq d \end{cases} \quad (3)$$

where E_1^L is the reflected wave amplitude, E_2^R and E_2^L correspond to the two counter-propagating waves in medium two, E_3^R is the transmitted wave amplitude, and the time-dependence $\exp(-i\omega t)$ has been omitted. We confine our attention to situations where medium three is passive, so that k_{3z}^R is uncontroversially the correct choice for the wavevector in medium three. The most direct route to solve for the four unknown wave amplitudes is to enforce Maxwell's boundary conditions: the transverse components of the E and H-fields must be continuous across the two interfaces at $z = 0$ and $z = d$, which results in four equations that can be solved for the four unknowns. The resulting reflection coefficient from the slab can be expressed in terms of the single-surface Fresnel coefficients as

$$r \equiv \frac{E_{1L}}{E_{1R}} = \frac{r_{12} + r_{23}\exp(2ik_{2z}^R d)}{1-\nu} \quad (4)$$

where

$$\nu = r_{21} r_{23} \exp(2ik_{2z}^R d) \quad (5)$$

is referred to as the roundtrip coefficient; the amplitude of a planewave circulating in the slab is multiplied by this factor after each roundtrip in the absence of any sources outside the slab. (Although we explicitly discuss s-polarized light, our conclusions as well as Eqs. 4 and 5 hold for both polarization states.) It is essential to note that the reflection coefficient given by Eq. 4 is invariant under the transformation $k_{2z}^R \leftrightarrows k_{2z}^L$; this is not surprising as it can be interpreted simply as a relabeling of the waves $E_2^R \leftrightarrows E_2^L$ in Eq. 3 that does not affect the final result. This invariance is important because it means that Eq. 4–which gives the reflection from the slab–is correct even if our postulate about the correct choice for k_{2z} in the single-surface problem turns out to be incorrect. We emphasize that the reflection coefficient given by Eq. 4 is a valid solution to Maxwell's equations for all material parameters, and in particular for any value of ν. The roundtrip coefficient ν has an important physical meaning, and intuitively one would expect three different regimes of behavior when the magnitude of ν is less than, equal to, or greater than one. The case where $|\nu| < 1$ governs passive slabs (in most but not all cases) and sufficiently weakly amplifying slabs. When $\nu = 1$ the slab behaves as a laser and emits light even in the absence of an incident wave, which manifests itself mathematically as an infinitely large reflection amplitude. The case where $|\nu| > 1$, however, has received scant attention in the literature [10].

Perhaps the reason for the neglect of the $|\nu| > 1$ steady-state solution is the seemingly intuitive assumption that there cannot be a steady-state solution when $|\nu| > 1$ (due to gain clamping in a laser, for instance). This assumption is only reinforced by examining a second well-known solution method for the reflection coefficient that decomposes the reflected wave amplitude E_1^L into a sum over partial waves, yielding the reflection coefficient

$$r = r_{12} + t_{12} t_{21} r_{23} \exp(2ik_{2z}^R d) \sum_{m=0}^{\infty} \nu^m. \quad (6)$$

Heuristically, the first term r_{12} (hereinafter referred to as the "specular" partial wave) of Eq. 6 results from the single-surface reflection of the incident wave at the 1-2 interface, and the geometric series accounts for the contributions to the reflected wave following multiple roundtrips within the slab. When $|\nu| < 1$, the geometric series in Eq. 6 converges to $(1-\nu)^{-1}$, giving the same result as found by matching the boundary conditions in Eq. 4. When $|\nu| > 1$, however, the geometric series diverges and the reflection coefficient is infinite. Intuitively, this divergence seems reasonable, since we expect any light that couples into a slab with $|\nu| > 1$ to be amplified after each roundtrip, and therefore grow without bound. Nevertheless, Eq. 4 yields a finite reflection coefficient even when $|\nu| > 1$, so how can we reconcile these two very different solutions?

To understand the non-divergent solution, we examine the behavior of a "finite-diameter" beam of light incident on the slab by numerically superposing the planewave solutions of Eq. 3, where E_1^L, E_2^R, E_2^L, and E_3^R are all determined by the convergent method of matching the boundary conditions. Let us consider the case where $\epsilon_1 = \epsilon_3 = 2.25$, the slab is an amplifying medium with $\epsilon_2 = 1 - 0.01i$, and $\mu_1 = \mu_2 = \mu_3 = 1$. We superpose a finite number of planewaves with incident angles in the range $27.47° < \theta < 32.53°$ and with amplitudes appropriate to generate a Gaussian (to within the sampling accuracy) beam incident on the slab at 30° with a full-width at half-maximum beam-diameter of 13.3 μm. The free-space wavelength of the beam is $\lambda_o = 1\ \mu$m. We can examine the transition at $|\nu| = 1$ simply by varying d, since both $|r_{21}|$ and $|r_{23}|$ are less than one (and independent of d), whereas $|\exp(2ik_{2z}^R d)|$ (and hence ν) increases monotonically with d (because k_{2z}^R has a negative imaginary part). A plot of the field $E_y(x,z)$ at one instant of time is shown in Fig. 2(a) for $d = 19\ \mu$m, which was chosen so that $|\nu|$ is slightly less than one for all constituent planewaves of the beam $(0.46 < |\nu| < 0.99)$. The arrows overlying the plot point in the direction of the time-averaged Poynting vector within their vicinity, indicating the direction of energy flow in the system, and the incident beam is uniquely identified by the white arrow. The beam behaves as we expect it to: the incident beam strikes the slab near $(x = 0,\ z = 0)$, giving rise to a specularly reflected beam as well as a refracted beam that 'zig-zags' up the slab, which in turn generates a

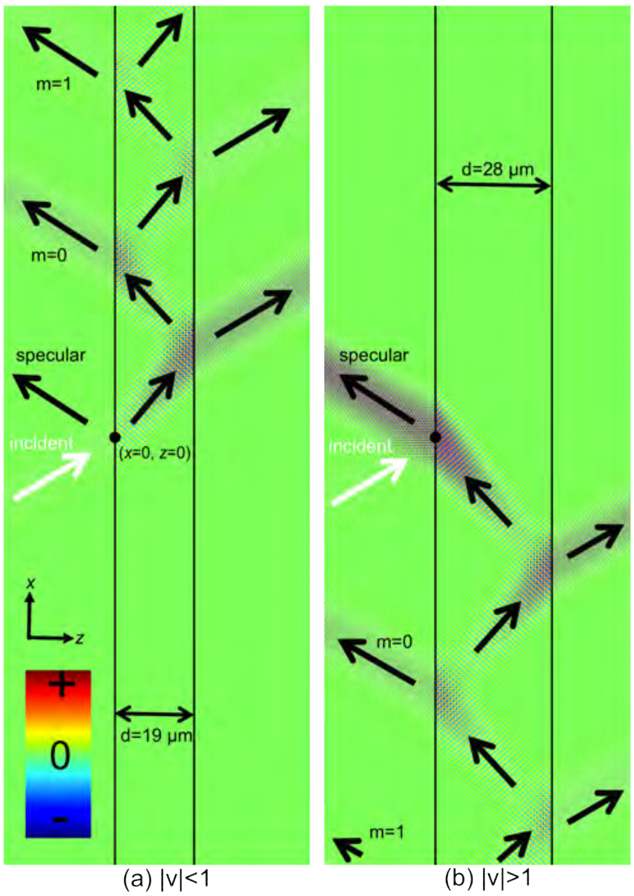

FIG. 2. Plots of the field $E_y(x, z)$ at one instant of time for a Gaussian beam (indicated with the white arrow) incident on an amplifying slab for (a) $d = 19$ μm and (b) $d = 28$ μm. Each reflected beam can be associated with a term in the appropriate partial wave sum, either Eq. 6 for (a) or Eq. 7 for (b). The black dot indicates the origin of the coordinate system.

reflected beam in medium one each time it strikes the 2-1 interface. (The field amplitude is plotted on a linear scale, and so the incident beam as well as the specularly reflected beam appear faint relative to the subsequently amplified portions of the beam.) Each of these reflected beams can intuitively be associated with a term of the partial wave expansion of Eq. 6–either the specular term or the mth term of the geometric series.

In Fig. 2(b) all parameters are kept the same except the slab thickness is increased to $d = 28$ μm, resulting in $|\nu|$ greater than one for all constituent planewaves of the Gaussian beam ($1.01 < |\nu| < 2.58$). Based solely on the plot of the field amplitude and not on the direction of energy flow indicated by the arrows, it may appear that the incident beam strikes the interface and negatively refracts in the slab, then zig-zags downwards in the $-\hat{\boldsymbol{x}}$ direction, giving rise to many reflected beams in medium one (and transmitted beams in medium three) which emanate from points on the slab with $x < 0$. Such an explanation was offered for simulations similar to ours [7, 8] to attempt to justify negative refraction in an active, nonmagnetic medium. However, by analyzing the Poynting vector we see that the energy in the beam zig-zags *up* the slab, so this phenomenon is distinct from negative refraction, despite the similarity in the positions of the reflected and transmitted beams. (In the supplemental information, a video of the time-dependent behavior of a "finite-duration" pulse of light more clearly illustrates the direction of energy flow.) The presence of energy in the slab at $x < 0$ has a perfectly causal explanation when one considers that the Gaussian beam does not have a truly finite spatial width, but rather a rapidly decaying "side-tail" in the direction normal to the propagation direction. The side-tail is capable of injecting a small amount of energy into the slab at positions $x \ll 0$. When $|\nu| > 1$, light in the slab gains more during one roundtrip than it loses to transmission at both facets, and so this initially small amount of energy is amplified, resulting in the "pre-excited" field seen at $x < 0$ in Fig. 2(b), so-called because the excitation occurs before the central lobe of the incident beam arrives at the slab. The key point is that when $|\nu| > 1$, our intuition about the arrival time and arrival position of the beam (or pulse) misleads us because amplification by the slab acts on typically negligible field amplitudes to dramatically alter the character of the field. Importantly, we see in Fig. 2(b) that when the pre-excited beam meets the incident beam at ($x = 0$, $z = 0$), the interference is such that all the energy in the slab leaves with the specularly reflected beam. In hindsight, this is necessary for the field to not diverge, since any energy remaining in the slab at this point would continue to zig-zag up and grow without bound. Finally, we emphasize that the specularly reflected beam is amplified relative to the incident beam as a result of the energy it receives from the pre-excited field in the slab, a mechanism that does not occur when $|\nu| < 1$.

Although the partial wave method predicts a divergent reflection coefficient when $|\nu| > 1$, with one small modification this method in fact offers significant insight into the $|\nu| > 1$ case. Recall that the reflection coefficient of Eq. 4 is invariant under the exchange $k^R_{2z} \leftrightarrows k^L_{2z}$. Applying this same transformation to the partial wave sum of Eq. 6 [10], we can express the reflection coefficient as

$$r = r'_{12} + t'_{12}t'_{21}r'_{23}\exp(2ik^L_{2z}d)\sum_{m=0}^{\infty}\nu'^m, \qquad (7)$$

where the prime indicates the substitution $k^R_{2z} \rightarrow k^L_{2z}$. Because the new roundtrip coefficient, $\nu' = r'_{21}r'_{23}\exp(2ik^L_{2z}d)$, is equal to ν^{-1}, in cases where $|\nu| > 1$ the primed partial wave sum of Eq. 7 will converge to the reflection coefficient of Eq. 4. Therefore, each reflected beam in Fig. 2(b) can be associated either with the specular term r'_{12} or with the mth term of the primed partial wave expansion in Eq. 7. In particular, note that the amplitude of the specularly reflected beam is r_{12}

when $|\nu| < 1$, which discontinuously changes to r'_{12} when $|\nu| > 1$. Because $|r_{12}| < 1$ (in most cases of practical interest) and $r'_{12} = r_{12}^{-1}$, this is a mathematical explanation for why the specular beam is amplified relative to the incident beam only when $|\nu| > 1$. Physically, we have seen from Fig. 2(b) that this specular amplification is made possible by the pre-excitation, a mechanism which cannot occur when $|\nu| < 1$.

It is interesting to compare a lossy and an amplifying slab in the limit as $d \to \infty$. In a lossy slab (for which $\mathrm{Im}(k_{2z}^R) > 0$), the roundtrip coefficient $\nu \to 0$ as $d \to \infty$, and so the reflection coefficient r approaches the single-surface solution r_{12}, as expected, because the geometric series in Eq. 6 makes no contribution. In a gainy slab (for which $\mathrm{Im}(k_{2z}^R) < 0$), $\nu \to \infty$ as $d \to \infty$, but $\nu' \to 0$ and so we see from Eq. 7 that $r \to r'_{12}$ (which means that the field in the slab is dominated by the wavevector k_{2z}^L, i.e., $E_2^R/E_2^L \to 0$). The reason the limiting treatment of $d \to \infty$ for a gainy slab does not yield the proper single-surface reflection coefficient is that no matter how large one chooses to make d, the nonzero reflection r_{23} at the back-facet of the slab allows for the amplification of the pre-excited field; for $|\nu| \gg 1$, this results in the left-propagating wavevector k_{2z}^L dominating the behavior of the slab while the amplitude of the wave associated with k_{2z}^R diminishes substantially, and the reflection coefficient correspondingly approaches r'_{12}. Nevertheless, the right-propagating wave is essential in spite of its seemingly inconsequential amplitude, as it is responsible for generating the left-propagating wave by way of the back-facet reflection. In the case of two truly semi-infinite media (i.e., media one and two), the absence of a back-facet prevents any roundtrip amplification of the pre-excitation, so the only wavevector that exists in the transmission medium is k_{2z}^R and the single-surface reflection coefficient is correctly given by r_{12}, not r'_{12}.

So far we have examined the relevance of the roundtrip coefficient ν only through its monotonic dependence on d, but ν is also a function of the incidence angle θ. For the same parameters as those used in Fig. 2(b), ν increases monotonically with an increasing incidence angle θ for s-polarized light; in particular, $|\nu|$ exceeds one as long as $\theta > 27.43°$. (For p-polarized light, ν increases monotonically with θ only once θ_2, the angle of propagation in medium two, exceeds the Brewster angles at both the 2-1 and 2-3 interfaces). As θ approaches and surpasses the critical angle for TIR, $\theta_c = 41.8°$, $|\nu|$ quickly becomes extremely large due to the negatively increasing $\mathrm{Im}(k_{2z}^R)$. (For $\theta = 41°$, $|\nu| = 9.34 \cdot 10^3$, and for $\theta = 42°$, $|\nu| = 1.40 \cdot 10^{15}$.) Thus, TIR from a gainy slab is well within the regime $|\nu| >> 1$ (for any reasonable thickness d), which, as previously argued, results in a reflection coefficient r'_{12}, and therefore the specular beam is amplified. It has been argued extensively that such amplification of the reflected beam is also possible when the gainy medium is truly semi-infinite [9–12, 14]; in other words, that the incident wave directly excites the wave with wavevector k_{2z}^L in medium two, resulting in the single-surface reflection coefficient r'_{12}. (This conjecture is known as single-surface amplified TIR.) It seems to us that a more unified and consistent approach would be to understand the situation $\theta > \theta_c$ simply as one way to achieve very large $|\nu|$ in a cavity. This would then be comparable to the case of large d, for which we demonstrated in the previous paragraph that the existence of the left-propagating k_{2z}^L relies on the nonzero back-facet reflection r_{23} [13]. This suggests that k_{2z}^R is the correct choice for the transmitted wavevector in the single-surface problem, even in the case of TIR from an amplifying medium.

For potential future research directions into the pre-excitation mechanism and its consequences, see the supplementary information.

The pulse simulation was run on the Odyssey cluster supported by the Harvard FAS Research Computing Group. TSM is supported by an NSF Graduate Research Fellowship. We thank Alexey Belyanin for helpful discussions.

---

[1] J.B. Pendry, "Negative refraction makes a perfect lens," Phys. Rev. Lett. **85**, 3966-69 (2000).

[2] D.R. Smith, D. Schuring, M. Rosenbluth, S. Schultz, S.A. Ramakrishna, and J.B. Pendry, "Limitations on subdiffraction imaging with a negative refractive index slab," Appl. Phys. Lett. **82**, 1506-1508 (2003).

[3] N. Garcia and M. Nieto-Vesperinas. "Left-Handed Materials Do Not Make a Perfect Lens." Phys. Rev Lett. **88**, 20, 207403 (2002).

[4] Y. Chen, P. Fischer, and F.W. Wise. "Negative Refraction at Optical Frequencies in Nonmagnetic Two-Component Molecular Media." Phys. Rev. Lett. **95**, (2005).

[5] S.A. Ramakrishna. "Comment on 'Negative Refraction at Optical Frequencies in Nonmagnetic Two-Component Molecular Media.'" Phys. Rev. Lett. **98**, (2007).

[6] S.A. Ramakrishna and O.J.F. Martin. "Resolving the wave vector in negative index media." Opt. Lett. **30**, 19, 2626-2628 (2005).

[7] J.B. Geddes III, T.G. Mackay, and A. Lakhtakia. "On the refractive index for a nonmagnetic two-component medium: Resolution of a controversy." Opt. Comm. **280**, 120-125 (2007).

[8] B. Nistad and J. Skaar. "Causality and elecromagnetic properties of active media." Phys. Rev. E **78**, 036603 (2008).

[9] C.J. Koester, "Laser action by enhanced total internal reflection," IEEE J. Quantum Electron. **QE-2**, 580-4 (1966)

[10] P.R. Callary and C.K. Carniglia. "Internal reflection from an amplifying layer." J. Opt. Sci. Am. **66**, 8 (1976).

[11] J. Fan, A. Dogariu, and L.J. Wang, "Amplified total internal reflection," Opt. Express **11**, 299-308 (2003).

[12] K.J. Willis, J.B. Schneider, and S.C. Hagness. "Amplified

total internal reflection: theory, analysis, and demonstration of existence via FDTD," Opt. Express **16**, 1903-14 (2008).
[13] A. Siegman, "Fresnel Reflection and Evanescent Gain," OPN, 38–45 (Jan. 2010).
[14] J.O. Grepstad and J. Skaar. "Total internal reflection and evanescent gain." Opt. Express **19**, 21404 (2011).

Fresnel reflection from a cavity with net roundtrip gain: Supplementary Information

Tobias S. Mansuripur[1,*] and Masud Mansuripur[2]

[1]*Department of Physics, Harvard University, Cambridge, MA 02139*
[2]*College of Optical Sciences, The University of Arizona, Tucson, AZ 85721*

I. PRESCRIPTION FOR R AND L SUPERSCRIPTS

The energy flux of an s-polarized planewave whose E-field is given by

$$E_y(x, z, t) = E_0 \exp(ik_x x + ik_z z - i\omega t) \tag{1}$$

in a medium (ϵ, μ) is given by the time-average of the Poynting vector $\boldsymbol{S} \equiv \vec{E} \times \vec{H}$,

$$\langle \vec{S} \rangle = \frac{|E_0|^2}{2\omega\mu_0} e^{-2\mathrm{Im}(k_z)z} \left(\mathrm{Re}\left[\frac{k_x}{\mu}\right] \hat{x} + \mathrm{Re}\left[\frac{k_z}{\mu}\right] \hat{z} \right). \tag{2}$$

Therefore, energy flows in the $+z$-direction ('to the right,' in our convention) when $\mathrm{Re}(k_z/\mu) > 0$, and we denote the wavevector k_z which satisfies this condition with the superscript R. Any value k_z for which $\mathrm{Re}(k_z/\mu) < 0$ is accordingly labeled with a superscript L. It follows from these definitions that k_z^R must make an acute angle with μ in the complex plane. (For p-polarized light energy flows in the $+z$-direction when $\mathrm{Re}(k_z/\epsilon) > 0$, and so k_z^R makes an acute angle with ϵ in the complex plane.)

In cases where $\langle S_z \rangle = 0$, we must establish a prescription for resolving the ambiguity in the choice of superscript, which is best illustrated by an example. Consider the case where medium one is a lossless dielectric ($\epsilon_1 > 1$, $\mu_1 = 1$), medium two is vacuum, and the incident propagating wave satisfies $k_x > k_0$, where $k_0 \equiv \omega/c$, so that the two choices for k_{2z} are $\pm i\sqrt{k_x^2 - k_0^2}$. Both choices for k_{2z} yield pure evanescent waves and carry no energy along the z-direction. By adding a small amount of loss to medium two, so that $\epsilon_2 = 1 + i\epsilon_2''$ where $\epsilon_2'' > 0$, the two choices for k_{2z} deviate slightly from the imaginary axis as shown in Supp. Fig. 1(a). Now both waves carry non-zero energy along the z-direction; the first quadrant solution is k_{2z}^R (which can be seen quickly because it makes an acute angle with $\mu_2 = 1$) and the third quadrant solution is k_{2z}^L. Our prescription to establish k_{2z}^R for a true vacuum (i.e., $\epsilon_2'' = 0$) is to take the limit $\epsilon_2'' \to 0$, which yields k_{2z}^R as the solution along the positive imaginary axis.

Beware that if one adds a small amount of gain rather than loss to medium two, so that $\epsilon_2 = 1 + i\epsilon_2''$ where $\epsilon_2'' < 0$, then the two solutions for k_{2z} exist in the second and fourth quadrants as shown in Supp. Fig. 1(b), and in this case k_{2z}^R points predominantly along the *negative* imaginary axis. Thus, we see that the two limiting cases as gain or loss approaches zero do not yield the same result:

$$\lim_{\epsilon_2'' \to 0^+} k_{2z}^R = - \lim_{\epsilon_2'' \to 0^-} k_{2z}^R. \tag{3}$$

To have an unambiguous labeling convention for the case $\epsilon_2'' = 0$, we emphasize that one must take the limit as *loss* approaches zero, which can be different from the limit as gain approaches zero in the case of evanescent waves.

Finally, it is worth noting that this discontinuity in the two limiting cases, apart from being a footnote in establishing a labeling convention, is actually at the heart of the debate over single-surface amplified TIR. When medium two has gain, if one chooses k_{2z}^R as the transmitted wavevector (in accordance with our postulate), then it seems unphysical that as the gain approaches (but does not reach) zero the transmitted wave should still be strongly amplified. To remedy this situation it has been suggested that the correct choice for the transmitted wavevector should be k_{2z}^L when medium two has gain and $k_x > k_0$, so that the transmitted wave decays in the $+z$-direction. We believe, instead, that the discontinuity in the two limits is not as unphysical as it might appear at first: the transmitted wave propagates a large distance in the x-direction while barely moving forward

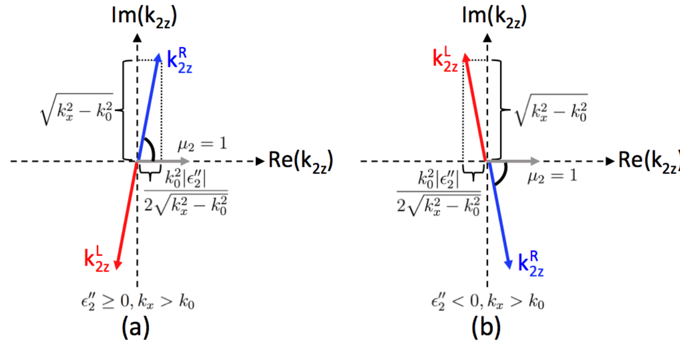

FIG. 1. Choosing the R or L label for an evanescent wave. (a) The two choices for k_{2z} are shown for the case of a slightly "lossy vacuum" ($\epsilon_2 = 1 + i\epsilon_2''$ where $\epsilon_2'' > 0$, $\mu_2 = 1$), for the case $k_x > k_0$. The first quadrant solution carries energy to the right and is labeled k_{2z}^R, and our prescription is to take the limit $\epsilon_2'' \to 0$ to determine that k_{2z}^R in the lossless case is along the positive imaginary axis. (b) For a slightly "gainy vacuum" ($\epsilon_2'' < 0$) the two solutions for k_{2z} are in the second and fourth quadrants, and k_{2z}^R approaches the negative imaginary axis as $\epsilon_2'' \to 0$. The magnitudes of the real and imaginary parts of k_{2z} in (a) and (b) are approximated using the first order Taylor expansion for small ϵ_2'': $k_0^2|\epsilon_2''| \ll k_x^2 - k_0^2$.

* mansuripur@physics.harvard.edu

in the z-direction (since $k_x \gg \mathrm{Re}(k_{2z}^R)$), so the large gain in the z-direction is actually a result of the long propagation distance along the x-direction. Far more unphysical, in our opinion, is the decision to switch the transmitted wavevector from k_{2z}^R when $k_x < k_0$ to k_{2z}^L when $k_x > k_0$. All of these arguments aside, however, the purpose of our paper has been to demonstrate a mechanism by which the specularly reflected beam from a finite-thickness slab can be amplified, both below and above the critical angle.

II. PULSE OF LIGHT INCIDENT ON GAINY SLAB WITH $|\nu| > 1$

The video file pulse_video.avi included online is a time-lapse video of $|E_y|^2$ of a pulse of light, rather than a beam, for the same material parameters as in Fig. 2(b) of the main text: $\epsilon_1 = \epsilon_3 = 2.25$, $\epsilon_2 = 1 - 0.01i$, $\mu_1 = \mu_2 = \mu_3 = 1$, $d = 28$ μm. The white vertical lines in the video identify the 1-2 and 2-3 interfaces. The incident pulse is s-polarized and Gaussian in both space (FWHM = 13.3 μm) and time (FWHM = 50 fs, or 15 optical cycles). The central wavelength of the pulse is $\lambda_o = 1$ μm, and the mean incidence angle (i.e., averaged over all constituent planewaves) is 30°. The size of each video frame is 210 μm by 150 μm (height by width). The time elapsed between frames is 10 fs, and the entire video spans 1.22 ps (123 frames total). The field $|E_y|^2$ is plotted on a logarithmic scale covering 3 decades, i.e. red corresponds to the maximum intensity and blue corresponds to intensities less than or equal to 1/1000th of the maximum. The background in this image is blue, which corresponds to the minimum of $|E_y|^2$, whereas the background in Figs. 2(a) and 2(b) is green because it is the field E_y that was plotted in that case, so that blue corresponded to the maximum negative field.

In the video, one first sees the incident pulse near the bottom left of the screen, traveling up and to the right. The pre-excitation is soon seen in the slab at the bottom of the frame, and the reflected pulse that corresponds to the $m = 1$ term in the primed partial wave expansion leaves the slab and propagates up and to the left in medium one. The pre-excitation in the slab then undergoes one roundtrip as it zig-zags upward, giving rise to a transmitted pulse in medium three followed by the $m = 0$ reflected pulse in medium one. The pre-excitation then makes one more roundtrip, giving rise to another transmitted pulse in medium three, and then approaches the 2-1 interface at the same time the incident pulse arrives from the other side. The two pulses interfere in such a way as to yield an amplified specularly reflected pulse by entirely depleting the energy content of the slab. The fact that the pre-excitation in the slab travels in the $+x$-direction clearly distinguishes this behavior from negative refraction.

III. FUTURE WORK

Our intent in this Letter has been to demonstrate the unintuitive behavior of a beam incident on a slab whose roundtrip coefficient is greater than one, which we demonstrated using analytical solutions to Maxwell's equations. We argued that when $|\nu| > 1$, the amplification of typically negligible field amplitudes results in a "pre-excited" beam in the slab which interferes with the incident beam to prevent the divergence of the field. We focused on two peculiar consequences of this phenomenon: 1) the specularly reflected beam is amplified only when $|\nu| > 1$, and 2) the field in the slab is dominated by the wavevector k_{2z}^L when $|\nu| \gg 1$. Our analysis has been restricted to the case of planar media with infinite extent in the x and y-dimensions, with homogeneous and frequency-independent material parameters. We do not consider this a serious drawback of our argument, as the majority of analyses of the three-layer problem employ the same assumptions. From a purely theoretical viewpoint and within the confines of the assumptions we have made, therefore, we believe that these results provide perspective for single-surface amplified TIR, and counter the notion of negative refraction in a nonmagnetic slab. Because the pre-excitation mechanism can only occur in a finite-thickness slab, we speculate that k_{2z}^R is the correct choice for the transmitted wavevector in the single-surface problem in all cases; in other words, the transmitted wave always carries energy away from the interface. There remain many open questions to be answered theoretically. Finite-difference time-domain simulations may be best suited to determine how the pre-excitation mechanism, which begins at $x \ll 0$, is affected when the slab has a finite length in the x-direction. Spontaneous emission and gain saturation must be accounted for in real materials. How would the slab behave if the beam had a truly finite width and no side-tail? To investigate how the slab reaches the steady state over time in response to a pulse with a sharp turn-on, the material parameters of the slab must be made to obey the Kramers-Kronig relations. Despite our lack of answers to these important questions, we hope that our analysis has clarified at least some of the relevant issues.

Surprisingly, $|\nu|$ can exceed one even in passive media provided $|r_{21}|$ or $|r_{23}|$ exceeds one, which can happen for incident evanescent waves near surface plasmon resonances. Understanding the role of ν in these cases can yield additional insight, particularly to the case of Pendry's lens [1, 2], and will be treated in future work.

First Report of Referee A, received May 29th, 2013:

In their work "Fresnel reflection from a cavity with net roundtrip gain" the authors consider a simple problem of transmission of the light pulse through a finite-sized slab of gainy media. The authors argue that contrary to the conventional wisdom, the "tail" of the excitation pulse will pre-excite the gainy material, followed by perfect cancellation of the light inside the gainy medium at the point of specular reflection.

The authors justify their arguments with numerical solutions of Maxwell equations. Although the exact techniques used by the authors are unclear, it appears that the authors use Fourier-expansion technique to calculate propagation of the pulse. The results of the calculations clearly support the claims by the authors.

I disagree with the results of this work. The use any technique that assumes plane-wave solution of Maxwell equations for situations with substantial gain is fundamentally flawed. The material with substantial gain is by definition affected by gain saturation, and is therefore nonlinear. Plane wave solutions, on the other hand, are linear in their nature.

Furthermore, a common artifact of all Fourier-expansion techniques is the fact that the resulting solution is essentially a Fourier series, not a Fourier integral. As such, the resulting solution is always periodic in space and time.

It is reasonable to believe that the results presented in this manuscript are the combination of Fourer-series artifacts and the implicit use of linear "steady-state" solutions of Maxwell equations in implicitly nonlinear system.

I therefore feel that this work should not be published in the present form.

One possible solution that may give authors a more convincing case is if they try using time-domain solutions of Maxwell equations (such as FDTD), where all excitation pulses are finite in space and time. I would gladly re-evaluate this work if such solutions support the claims of this manuscript, provided of course, that applicability of the particular time-domain solver for materials with strong gain is well-established.

Response of authors, sent June 4th, 2013:

Dear Editor and Referee:

The referee's concerns that gain saturation and Fourier series artifacts in the field plots compromise the validity of our results are valid concerns, and we appreciate that the referee has grasped the problem deeply enough to point out these issues. We have added two sections to the Supplementary Information of the manuscript, one section on gain saturation and another section that provides the specifics of the Fourier series calculation method along with evidence that our results are not compromised by numerical artifacts. We would greatly appreciate if the referee were given an opportunity to read these two new sections we have added to address his/her concerns, in the hope that he/she may reconsider the validity of our results. Below, we address the referee's specific concerns and provide additional context for our arguments that is not included in the Supplementary Information.

Referee: "The use of any technique that assumes plane-wave solution of Maxwell equations for situations with substantial gain is fundamentally flawed. The material with substantial gain is by definition affected by gain saturation, and is therefore nonlinear. Plane wave solutions, on the other hand, are linear in their nature."

Response: Gain saturation is an effect that depends on the intensity of the light within the gain medium, not—as the referee contends—on the presence of "substantial" gain within the medium, which is not a precise criterion. Consider a wave with intensity I_0 which enters a single-pass traveling-wave amplifier that has gain g per unit length: the light intensity I within the amplifier will be given by $I=I_0\exp(gz)$, where z is the distance along the amplifier, until the light intensity becomes large enough to cause the gain g to saturate. Therefore, saturation is a function of I, not of g. In the Supplementary Information, we explain that, in the case of a laser, the referee's comment is valid: any value of gain above threshold will cause the light intensity to be amplified to as large a value as needed until gain saturation kicks in and reduces the gain. However, the problem considered in our manuscript is different because there is a field incident on the gain medium that prevents the field within the gain medium from diverging. The very fact that there exists a linear solution for which the field amplitude does not diverge (namely, the solution we discuss in the manuscript) means that it is possible to arbitrarily reduce the intensity of the incident field and avoid the complications of gain saturation.

Furthermore, we would like to emphasize that the method of using planewaves with gain media is not new to this manuscript. Many of the cited references that deal with the problem of single-surface amplified total internal reflection (ss-aTIR), or with the problem of negative refraction in gainy media, employ the same methodology. In fact, we pointed out in our manuscript that simulations similar to ours in other papers even achieved the same field plots. The main difference is that, by analyzing the Poynting vector in the slab, we are able to provide a different interpretation of the result: namely, that the pre-excitation, rather than negative refraction, causes the observed field distribution.

Referee: "Furthermore, a common artifact of all Fourier-expansion techniques is the fact that the resulting solution is essentially a Fourier series, not a Fourier integral. As such, the resulting solution is always periodic in space and time. It is reasonable to believe that the results presented in this manuscript are the combination of Fourier-series artifacts and the implicit use of linear "steady-state" solutions of Maxwell equations in implicitly nonlinear system."

Response: The referee is correct in assuming that our calculation method employs a Fourier series technique. We added a section to the Supplementary Information that provides the details of this numerical method. We also performed checks against numerical artifacts by varying the sampling rate in the Fourier domain, thereby changing the spatial periodicity of the infinite number of beams that arrive at the slab's surface. (Note that the periodicity introduced by discretization is several thousand microns along the x-axis, which is well outside the range of the plots in Fig. 2.) Regardless of the specific sampling rate used (so long as the rate is large enough), we have confirmed that the simulations yield the same results, and we are confident that our plots in Fig. 2 of the manuscript faithfully represent the behavior of the slab in response to only a single incident beam (not a periodic array of beams). Although it was important to perform these checks, the fact that the primed partial wave expansion in Eq. 7 correctly predicts the amplitudes of all the reflected and transmitted beams (in the case where the roundtrip coefficient exceeds one) is a strong argument against the notion that numerical artifacts are playing any role in our simulations. The primed partial wave expansion should not be overlooked, because it provides the mathematical description of the counter-intuitive behavior of the beam when the roundtrip coefficient exceeds one.

Referee: "One possible solution that may give authors a more convincing case is if they try using time-domain solutions of Maxwell equations (such as FDTD), where all excitation pulses are finite in space and time. I would gladly re-evaluate this work if such solutions support the claims of this manuscript, provided of course, that applicability of the particular time-domain solver for materials with strong gain is well-established."

Response: We agree that FDTD calculations would be nice to confirm the results of this work, but we would like to point out that FDTD simulations of amplifying media come with a few complications. The most significant one we have encountered is that the perfectly matched layer (PML) typically used to define an absorbing boundary around the simulation region has a very small but non-zero reflectivity. While this reflectivity is too small to appreciably affect simulations of passive media, a gain medium can amplify this initially weak reflected wave to large values, and so the boundary region of the FDTD simulation introduces its own numerical artifacts. Reference 12 [Willis] applies FDTD methods to the ss-aTIR problem and reaches the conclusion that the wave reflected by TIR from a semi-infinite gain medium can be amplified, but we would argue that the problem of the imperfect PML was not paid due attention. The results of our manuscript show that even a residual PML reflectivity on the order of roughly 10^{-30} (the exact value depends on the specific parameters of the problem) would allow the roundtrip coefficient to exceed one, simply because the evanescent wave gain $|\exp(2ik_{2z}^{R}d)|$ within the slab is enormous. In other words, the semi-infinite transmission medium the FDTD simulations in [12] were intended to simulate was not semi-infinite, and the amplified reflection can instead likely be understood in terms of reflection from a slab with roundtrip coefficient greater than one.

We had mentioned in the final section of the Supplementary Information that FDTD simulations could be useful to see how this pre-excitation phenomenon is affected when the slab is given a finite length in the *x*-direction, so that planewaves are no longer the eigenfunctions of Maxwell's equations. However, the nonzero reflectivity of the PML at the *x*-facets will need to be carefully accounted for, because the small reflections at these boundaries—when coupled into a slab with roundtrip coefficient greater than one—could easily cause field divergences that are only artifacts of the residual PML reflectivity.

Finally, we wish to emphasize that we have used our results to comment on the problem of ss-aTIR, which has been discussed in the literature for over 40 years. The essence of this problem is simple to state: when a *planewave* is incident on a gain medium above the critical angle, which wavevector must one choose for the transmitted *planewave*? The problem is fundamentally about *planewaves*. Therefore, while FDTD results are certainly interesting and could shed light on different aspects of the problem, the most convincing solution of the problem should be one that is ultimately derived from considerations of planewaves. By and large, this is the type of problem that our manuscript has tried to elucidate.

Thank you for your continued consideration of our manuscript.

Tobias Mansuripur

Version 2 of the manuscript (submitted to PRL 6/4/2013)

Fresnel reflection from a cavity with net roundtrip gain

Tobias S. Mansuripur[1,*] and Masud Mansuripur[2]

[1]Department of Physics, Harvard University, Cambridge, MA 02139
[2]College of Optical Sciences, The University of Arizona, Tucson, AZ 85721

A planewave incident on an active etalon with net roundtrip gain may be expected to diverge in field amplitude, yet Maxwell's equations admit only a convergent solution. By examining a Gaussian beam obliquely incident on such a cavity, we find that the "side-tail" of the beam leaks into the cavity and gives rise to a field that interferes with the main portion of the beam, which is ultimately responsible for the convergence of the field. This mechanism offers perspective for many phenomena, and we specifically discuss the implications for amplified total internal reflection.

The Fresnel coefficients govern the reflection and transmission of light for the simplest possible scenarios: at planar interfaces between homogeneous media. Despite this simplicity, some interesting solutions have been discovered only recently, such as 1) the amplification of evanescent waves in a passive, negative-index slab [1–3], and controversy regarding the proper choice of the wavevector in active media has persisted in relation to the possibility of 2) negative refraction in nonmagnetic media [4–8] as well as 3) single-surface amplified total internal reflection (TIR) [9–14]. It turns out that all three of these cases share a common thread: the presence of a cavity whose roundtrip gain exceeds the loss. In this Letter, we explore more generally the response of such a cavity to an incident beam of light.

To begin, we establish conventions that allow us to more clearly discuss the direction of energy flow, which is central to our overall argument. For the single-surface problem, shown in Fig. 1(a), the incident wavevector in medium one is $\boldsymbol{k}_1^R = k_x\hat{\boldsymbol{x}} + k_{1z}^R\hat{\boldsymbol{z}}$, and the reflected wavevector is $\boldsymbol{k}_1^L = k_x\hat{\boldsymbol{x}} + k_{1z}^L\hat{\boldsymbol{z}}$, where $k_{1z}^L = -k_{1z}^R$. The component k_x (which we assume for simplicity to be real-valued), once determined by the incident wave, must become the x-component of every wavevector in the system in order to satisfy Maxwell's boundary conditions. For the transmitted wavevector $\boldsymbol{k}_2$, the dispersion relation offers two choices for k_{2z},

$$k_{2z} = \pm\sqrt{(\omega/c)^2\mu_2\epsilon_2 - k_x^2}, \tag{1}$$

where ω is the angular frequency of the planewave and c is the speed of light in vacuum. We denote by k_{2z}^R (k_{2z}^L) the choice which carries energy to the right (left), namely, that for which the time-averaged z-component of the Poynting vector is positive (negative). (In cases where both choices for k_{2z} result in no energy flow in the z-direction, such as for evanescent waves in a transparent medium, our prescription is to add a small amount of loss to the slab which will unambiguously distinguish k_{2z}^R and k_{2z}^L, then take the limit as the loss goes to zero. See the supplemental information for details.) Let us

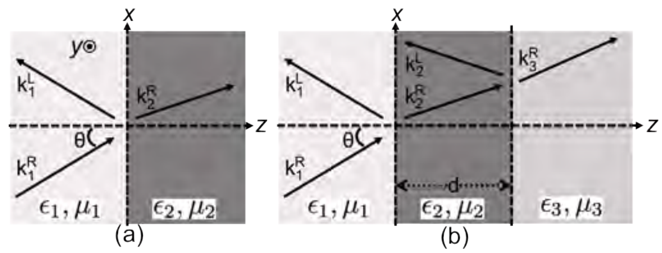

FIG. 1. Geometry of the (a) single-surface and (b) cavity problems. All media are infinite in the x and y-directions. The arrows denote the wavevectors of the planewaves present in each layer. The material constants are the relative permittivities and permeabilities.

postulate that the proper choice for k_{2z} in the single-surface problem is always k_{2z}^R (i.e., that the transmitted energy flows away from the interface), irrespective of the material parameters or the nature of the incident wave. (In fact, k_{2z}^R is universally agreed to be the correct choice in all cases except possibly that of amplified TIR; it is due to this one controversy that we refer to this choice as a postulate for now.) We require this postulate in order to unambiguously define the single-surface Fresnel reflection and transmission coefficients

$$r_{\ell m} = \frac{\tilde{k}_{\ell z}^R - \tilde{k}_{mz}^R}{\tilde{k}_{\ell z}^R + \tilde{k}_{mz}^R}, \; t_{\ell m} = \frac{2\tilde{k}_{\ell z}^R}{\tilde{k}_{\ell z}^R + \tilde{k}_{mz}^R} \tag{2}$$

where we have generalized the result for incidence medium ℓ and transmission medium m. For s-polarization we have defined $\tilde{k}_{nz} \equiv k_{nz}/\mu_n$, and the two coefficients yield the reflected and transmitted amplitudes of the component E_y, while for p-polarization $\tilde{k}_{nz} \equiv k_{nz}/\epsilon_n$ and the coefficients are associated with the component E_x.

Having established these conventions, we now consider the case of light incident on a cavity, shown in Fig. 1(b). The total E-field resulting from an s-polarized incident

* mansuripur@physics.harvard.edu

wave in medium one with amplitude E_1^R is given by

$$E_y(x,z) = \begin{cases} E_1^R \exp(ik_x x + ik_{1z}^R z) \\ \quad +E_1^L \exp(ik_x x + ik_{1z}^L z) & : z \leq 0 \\ E_2^R \exp(ik_x x + ik_{2z}^R z) \\ \quad +E_2^L \exp(ik_x x + ik_{2z}^L z) & : 0 \leq z \leq d \\ E_3^R \exp[ik_x x + ik_{3z}^R (z-d)] & : z \geq d \end{cases} \tag{3}$$

where E_1^L is the reflected wave amplitude, E_2^R and E_2^L correspond to the two counter-propagating waves in medium two, E_3^R is the transmitted wave amplitude, and the time-dependence $\exp(-i\omega t)$ has been omitted. We confine our attention to situations where medium three is passive, so that k_{3z}^R is uncontroversially the correct choice for the wavevector in medium three. The most direct route to solve for the four unknown wave amplitudes is to enforce Maxwell's boundary conditions: the transverse components of the E and H-fields must be continuous across the two interfaces at $z = 0$ and $z = d$, which results in four equations that can be solved for the four unknowns. The resulting reflection coefficient from the slab can be expressed in terms of the single-surface Fresnel coefficients as

$$r \equiv \frac{E_1^L}{E_1^R} = \frac{r_{12} + r_{23}\exp(2ik_{2z}^R d)}{1-\nu} \tag{4}$$

where

$$\nu = r_{21} r_{23} \exp(2ik_{2z}^R d) \tag{5}$$

is referred to as the roundtrip coefficient; the amplitude of a planewave circulating in the slab is multiplied by this factor after each roundtrip in the absence of any sources outside the slab. (Although we explicitly discuss s-polarized light, our conclusions as well as Eqs. 4 and 5 hold for both polarization states.) It is essential to note that the reflection coefficient given by Eq. 4 is invariant under the transformation $k_{2z}^R \leftrightarrows k_{2z}^L$; this is not surprising as it can be interpreted simply as a relabeling of the waves $E_2^R \leftrightarrows E_2^L$ in Eq. 3 that does not affect the final result. This invariance is important because it means that Eq. 4–which gives the reflection from the slab–is correct even if our postulate about the correct choice for k_{2z} in the single-surface problem turns out to be incorrect. We emphasize that the reflection coefficient given by Eq. 4 is a valid solution to Maxwell's equations for all material parameters, and in particular for any value of ν. The roundtrip coefficient ν has an important physical meaning, and intuitively one would expect three different regimes of behavior when the magnitude of ν is less than, equal to, or greater than one. The case where $|\nu| < 1$ governs passive slabs (in most but not all cases) and sufficiently weakly amplifying slabs. When $\nu = 1$ the slab behaves as a laser and emits light even in the absence of an incident wave, which manifests itself mathematically as an infinitely large reflection amplitude. The case where $|\nu| > 1$, however, has received scant attention in the literature [10].

Perhaps the reason for the neglect of the $|\nu| > 1$ steady-state solution is the seemingly intuitive assumption that there cannot be a steady-state solution when $|\nu| > 1$ (due to gain saturation in a laser, for instance. See the supplemental information for details.) This assumption is only reinforced by examining a second well-known solution method for the reflection coefficient that decomposes the reflected wave amplitude E_1^L into a sum over partial waves, yielding the reflection coefficient

$$r = r_{12} + t_{12} t_{21} r_{23} \exp(2ik_{2z}^R d) \sum_{m=0}^{\infty} \nu^m. \tag{6}$$

Heuristically, the first term r_{12} (hereinafter referred to as the "specular" partial wave) of Eq. 6 results from the single-surface reflection of the incident wave at the 1-2 interface, and the geometric series accounts for the contributions to the reflected wave following multiple roundtrips within the slab. When $|\nu| < 1$, the geometric series in Eq. 6 converges to $(1-\nu)^{-1}$, giving the same result as found by matching the boundary conditions in Eq. 4. When $|\nu| > 1$, however, the geometric series diverges and the reflection coefficient is infinite. Intuitively, this divergence seems reasonable, since we expect any light that couples into a slab with $|\nu| > 1$ to be amplified after each roundtrip, and therefore grow without bound. Nevertheless, Eq. 4 yields a finite reflection coefficient even when $|\nu| > 1$, so how can we reconcile these two very different solutions?

To understand the non-divergent solution, we examine the behavior of a "finite-diameter" beam of light incident on the slab by numerically superposing the planewave solutions of Eq. 3, where E_1^L, E_2^R, E_2^L, and E_3^R are all determined by the convergent method of matching the boundary conditions (see supplementary information for details). Let us consider the case where $\epsilon_1 = \epsilon_3 = 2.25$, the slab is an amplifying medium with $\epsilon_2 = 1 - 0.01i$, and $\mu_1 = \mu_2 = \mu_3 = 1$. We superpose a finite number of planewaves with incident angles in the range $27.47° < \theta < 32.53°$ and with amplitudes appropriate to generate a Gaussian (to within the sampling accuracy) beam incident on the slab at 30° with a full-width at half-maximum beam-diameter of 13.3 μm. The free-space wavelength of the beam is $\lambda_o = 1$ μm. We can examine the transition at $|\nu| = 1$ simply by varying d, since both $|r_{21}|$ and $|r_{23}|$ are less than one (and independent of d), whereas $|\exp(2ik_{2z}^R d)|$ (and hence ν) increases monotonically with d (because k_{2z}^R has a negative imaginary part). A plot of the field $E_y(x,z)$ at one instant of time is shown in Fig. 2(a) for $d = 19$ μm, which was chosen so that $|\nu|$ is slightly less than one for all constituent planewaves of the beam $(0.46 < |\nu| < 0.99)$. The arrows overlying the plot point in the direction of the time-averaged Poynting vector within their vicinity, indicating the direction of energy flow in the system, and the incident beam is uniquely identified by the white arrow. The beam behaves as we expect it to: the incident beam strikes the slab near $(x = 0,\ z = 0)$, giving rise to

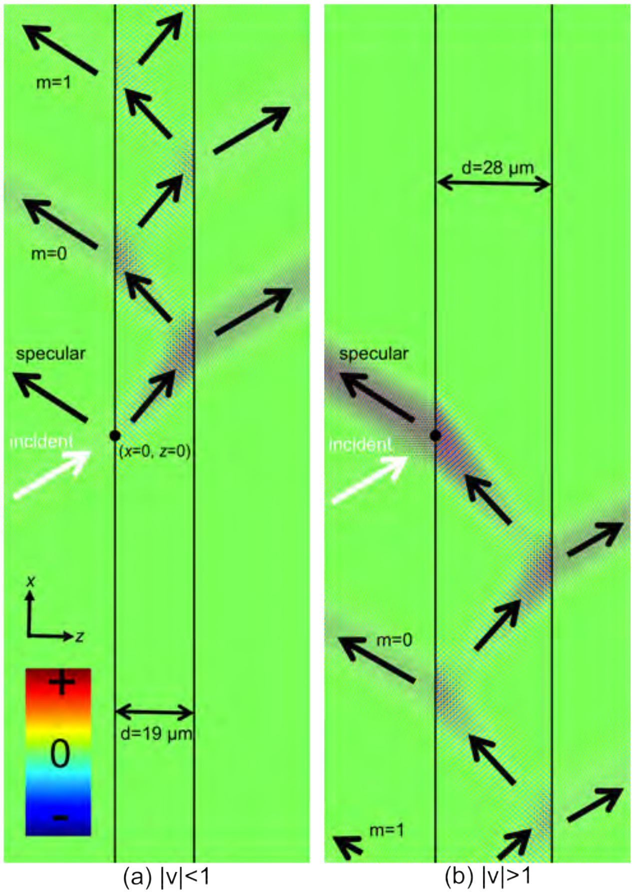

FIG. 2. Plots of the field $E_y(x, z)$ at one instant of time for a Gaussian beam (indicated with the white arrow) incident on an amplifying slab for (a) $d = 19$ μm and (b) $d = 28$ μm. Each reflected beam can be associated with a term in the appropriate partial wave sum, either Eq. 6 for (a) or Eq. 7 for (b). The black dot indicates the origin of the coordinate system.

a specularly reflected beam as well as a refracted beam that 'zig-zags' up the slab, which in turn generates a reflected beam in medium one each time it strikes the 2-1 interface. (The field amplitude is plotted on a linear scale, and so the incident beam as well as the specularly reflected beam appear faint relative to the subsequently amplified portions of the beam.) Each of these reflected beams can intuitively be associated with a term of the partial wave expansion of Eq. 6–either the specular term or the mth term of the geometric series.

In Fig. 2(b) all parameters are kept the same except the slab thickness is increased to $d = 28$ μm, resulting in $|\nu|$ greater than one for all constituent planewaves of the Gaussian beam ($1.01 < |\nu| < 2.58$). Based solely on the plot of the field amplitude and not on the direction of energy flow indicated by the arrows, it may appear that the incident beam strikes the interface and negatively refracts in the slab, then zig-zags downwards in the $-\hat{\boldsymbol{x}}$ direction, giving rise to many reflected beams in medium one (and transmitted beams in medium three) which emanate from points on the slab with $x < 0$. Such an explanation was offered for simulations similar to ours [7, 8] to attempt to justify negative refraction in an active, non-magnetic medium. However, by analyzing the Poynting vector we see that the energy in the beam zig-zags *up* the slab, so this phenomenon is distinct from negative refraction, despite the similarity in the positions of the reflected and transmitted beams. (In the supplemental information, a video of the time-dependent behavior of a "finite-duration" pulse of light more clearly illustrates the direction of energy flow.) The presence of energy in the slab at $x < 0$ has a perfectly causal explanation when one considers that the Gaussian beam does not have a truly finite spatial width, but rather a rapidly decaying "side-tail" in the direction normal to the propagation direction. The side-tail is capable of injecting a small amount of energy into the slab at positions $x \ll 0$. When $|\nu| > 1$, light in the slab gains more during one roundtrip than it loses to transmission at both facets, and so this initially small amount of energy is amplified, resulting in the "pre-excited" field seen at $x < 0$ in Fig. 2(b), so-called because the excitation occurs before the central lobe of the incident beam arrives at the slab. The key point is that when $|\nu| > 1$, our intuition about the arrival time and arrival position of the beam (or pulse) misleads us because amplification by the slab acts on typically negligible field amplitudes to dramatically alter the character of the field. Importantly, we see in Fig. 2(b) that when the pre-excited beam meets the incident beam at ($x = 0$, $z = 0$), the interference is such that all the energy in the slab leaves with the specularly reflected beam. In hindsight, this is necessary for the field to not diverge, since any energy remaining in the slab at this point would continue to zig-zag up and grow without bound. Finally, we emphasize that the specularly reflected beam is amplified relative to the incident beam as a result of the energy it receives from the pre-excited field in the slab, a mechanism that does not occur when $|\nu| < 1$.

Although the partial wave method predicts a divergent reflection coefficient when $|\nu| > 1$, with one small modification this method in fact offers significant insight into the $|\nu| > 1$ case. Recall that the reflection coefficient of Eq. 4 is invariant under the exchange $k^R_{2z} \leftrightharpoons k^L_{2z}$. Applying this same transformation to the partial wave sum of Eq. 6 [10], we can express the reflection coefficient as

$$r = r'_{12} + t'_{12}t'_{21}r'_{23}\exp(2ik^L_{2z}d)\sum_{m=0}^{\infty}\nu'^m, \qquad (7)$$

where the prime indicates the substitution $k^R_{2z} \rightarrow k^L_{2z}$. Because the new roundtrip coefficient, $\nu' = r'_{21}r'_{23}\exp(2ik^L_{2z}d)$, is equal to ν^{-1}, in cases where $|\nu| > 1$ the primed partial wave sum of Eq. 7 will converge to the reflection coefficient of Eq. 4. Therefore, each reflected beam in Fig. 2(b) can be associated either with the specular term r'_{12} or with the mth term of the primed partial wave expansion in Eq. 7. In particular, note that the amplitude of the specularly reflected beam is r_{12} when $|\nu| < 1$, which discontinuously changes to r'_{12} when $|\nu| > 1$. Because $|r_{12}| < 1$ (in most cases of practical interest) and $r'_{12} = r_{12}^{-1}$, this is a mathematical explanation for why the specular beam is amplified relative to the incident beam only when $|\nu| > 1$. Physically, we have seen from Fig. 2(b) that this specular amplification is made possible by the pre-excitation, a mechanism which cannot occur when $|\nu| < 1$.

It is interesting to compare a lossy and an amplifying slab in the limit as $d \to \infty$. In a lossy slab (for which $\mathrm{Im}(k_{2z}^R) > 0$), the roundtrip coefficient $\nu \to 0$ as $d \to \infty$, and so the reflection coefficient r approaches the single-surface solution r_{12}, as expected, because the geometric series in Eq. 6 makes no contribution. In a gainy slab (for which $\mathrm{Im}(k_{2z}^R) < 0$), $\nu \to \infty$ as $d \to \infty$, but $\nu' \to 0$ and so we see from Eq. 7 that $r \to r'_{12}$ (which means that the field in the slab is dominated by the wavevector k_{2z}^L, i.e., $E_2^R/E_2^L \to 0$). The reason the limiting treatment of $d \to \infty$ for a gainy slab does not yield the proper single-surface reflection coefficient is that no matter how large one chooses to make d, the nonzero reflection r_{23} at the back-facet of the slab allows for the amplification of the pre-excited field; for $|\nu| \gg 1$, this results in the left-propagating wavevector k_{2z}^L dominating the behavior of the slab while the amplitude of the wave associated with k_{2z}^R diminishes substantially, and the reflection coefficient correspondingly approaches r'_{12}. Nevertheless, the right-propagating wave is essential in spite of its seemingly inconsequential amplitude, as it is responsible for generating the left-propagating wave by way of the back-facet reflection. In the case of two truly semi-infinite media (i.e., media one and two), the absence of a back-facet prevents any roundtrip amplification of the pre-excitation, so the only wavevector that exists in the transmission medium is k_{2z}^R and the single-surface reflection coefficient is correctly given by r_{12}, not r'_{12}.

So far we have examined the relevance of the roundtrip coefficient ν only through its monotonic dependence on d, but ν is also a function of the incidence angle θ. For the same parameters as those used in Fig. 2(b), ν increases monotonically with an increasing incidence angle θ for s-polarized light; in particular, $|\nu|$ exceeds one as long as $\theta > 27.43°$. (For p-polarized light, ν increases monotonically with θ only once θ_2, the angle of propagation in medium two, exceeds the Brewster angles at both the 2-1 and 2-3 interfaces). As θ approaches and surpasses the critical angle for TIR, $\theta_c = 41.8°$, $|\nu|$ quickly becomes extremely large due to the negatively increasing $\mathrm{Im}(k_{2z}^R)$. (For $\theta = 41°$, $|\nu| = 9.34 \cdot 10^3$, and for $\theta = 42°$, $|\nu| = 1.40 \cdot 10^{15}$.) Thus, TIR from a gainy slab is well within the regime $|\nu| >> 1$ (for any reasonable thickness d), which, as previously argued, results in a reflection coefficient r'_{12}, and therefore the specular beam is amplified. It has been argued extensively that such amplification of the reflected beam is also possible when the gainy medium is truly semi-infinite [9–12, 14]; in other words, that the incident wave directly excites the wave with wavevector k_{2z}^L in medium two, resulting in the single-surface reflection coefficient r'_{12}. (This conjecture is known as single-surface amplified TIR.) It seems to us that a more unified and consistent approach would be to understand the situation $\theta > \theta_c$ simply as one way to achieve very large $|\nu|$ in a cavity. This would then be comparable to the case of large d, for which we demonstrated in the previous paragraph that the existence of the left-propagating k_{2z}^L relies on the nonzero back-facet reflection r_{23} [13]. This suggests that k_{2z}^R is the correct choice for the transmitted wavevector in the single-surface problem, even in the case of TIR from an amplifying medium.

For potential future research directions into the pre-excitation mechanism and its consequences, see the supplementary information.

The pulse simulation was run on the Odyssey cluster supported by the Harvard FAS Research Computing Group. TSM is supported by an NSF Graduate Research Fellowship. We thank Alexey Belyanin for helpful discussions.

---

[1] J.B. Pendry, "Negative refraction makes a perfect lens," Phys. Rev. Lett. **85**, 3966-69 (2000).

[2] D.R. Smith, D. Schuring, M. Rosenbluth, S. Schultz, S.A. Ramakrishna, and J.B. Pendry, "Limitations on subdiffraction imaging with a negative refractive index slab," Appl. Phys. Lett. **82**, 1506-1508 (2003).

[3] N. Garcia and M. Nieto-Vesperinas. "Left-Handed Materials Do Not Make a Perfect Lens." Phys. Rev Lett. **88**, 20, 207403 (2002).

[4] Y. Chen, P. Fischer, and F.W. Wise. "Negative Refraction at Optical Frequencies in Nonmagnetic Two-Component Molecular Media." Phys. Rev. Lett. **95**, (2005).

[5] S.A. Ramakrishna. "Comment on 'Negative Refraction at Optical Frequencies in Nonmagnetic Two-Component Molecular Media.'" Phys. Rev. Lett. **98**, (2007).

[6] S.A. Ramakrishna and O.J.F. Martin. "Resolving the wave vector in negative index media." Opt. Lett. **30**, 19, 2626-2628 (2005).

[7] J.B. Geddes III, T.G. Mackay, and A. Lakhtakia. "On the refractive index for a nonmagnetic two-component medium: Resolution of a controversy." Opt. Comm. **280**, 120-125 (2007).

[8] B. Nistad and J. Skaar. "Causality and elecromagnetic properties of active media." Phys. Rev. E **78**, 036603 (2008).

[9] C.J. Koester, "Laser action by enhanced total internal reflection," IEEE J. Quantum Electron. **QE-2**, 580-4 (1966)

[10] P.R. Callary and C.K. Carniglia. "Internal reflection from an amplifying layer." J. Opt. Sci. Am. **66**, 8 (1976).

[11] J. Fan, A. Dogariu, and L.J. Wang, "Amplified total in-

ternal reflection," Opt. Express **11**, 299-308 (2003).
[12] K.J. Willis, J.B. Schneider, and S.C. Hagness. "Amplified total internal reflection: theory, analysis, and demonstration of existence via FDTD," Opt. Express **16**, 1903-14 (2008).
[13] A. Siegman, "Fresnel Reflection and Evanescent Gain," OPN, 38–45 (Jan. 2010).
[14] J.O. Grepstad and J. Skaar. "Total internal reflection and evanescent gain." Opt. Express **19**, 21404 (2011).

Supplementary information for version 2 (submitted to PRL 6/4/2013).
Changes relative to version 1 are highlighted.

Fresnel reflection from a cavity with net roundtrip gain: Supplementary Information

Tobias S. Mansuripur[1,*] and Masud Mansuripur[2]

[1] *Department of Physics, Harvard University, Cambridge, MA 02139*

[2] *College of Optical Sciences, The University of Arizona, Tucson, AZ 85721*

[*] mansuripur@physics.harvard.edu

I. PRESCRIPTION FOR R AND L SUPERSCRIPTS

The energy flux of an s-polarized planewave whose E-field is given by

$$E_y(x, z, t) = E_0 \exp(ik_x x + ik_z z - i\omega t) \tag{1}$$

in a medium (ϵ, μ) is given by the time-average of the Poynting vector $\boldsymbol{S} \equiv \vec{E} \times \vec{H}$,

$$\langle \vec{S} \rangle = \frac{|E_0|^2}{2\omega\mu_0} e^{-2\text{Im}(k_z)z} \left(\text{Re}\left[\frac{k_x}{\mu}\right] \hat{x} + \text{Re}\left[\frac{k_z}{\mu}\right] \hat{z} \right). \tag{2}$$

Therefore, energy flows in the $+z$-direction ('to the right,' in our convention) when $\text{Re}(k_z/\mu) > 0$, and we denote the wavevector k_z which satisfies this condition with the superscript R. Any value k_z for which $\text{Re}(k_z/\mu) < 0$ is accordingly labeled with a superscript L. It follows from these definitions that k_z^R must make an acute angle with μ in the complex plane. (For p-polarized light energy flows in the $+z$-direction when $\text{Re}(k_z/\epsilon) > 0$, and so k_z^R makes an acute angle with ϵ in the complex plane.)

In cases where $\langle S_z \rangle = 0$, we must establish a prescription for resolving the ambiguity in the choice of superscript, which is best illustrated by an example. Consider the case where medium one is a lossless dielectric ($\epsilon_1 > 1$, $\mu_1 = 1$), medium two is vacuum, and the incident propagating wave satisfies $k_x > k_0$, where $k_0 \equiv \omega/c$, so that the two choices for k_{2z} are $\pm i\sqrt{k_x^2 - k_0^2}$. Both choices for k_{2z} yield pure evanescent waves and carry no energy along the z-direction. By adding a small amount of loss to medium two, so that $\epsilon_2 = 1 + i\epsilon_2''$ where $\epsilon_2'' > 0$, the two choices for k_{2z} deviate slightly from the imaginary axis as shown in Supp. Fig. 1(a). Now both waves carry non-zero energy along the z-direction; the first quadrant solution is k_{2z}^R (which can be seen quickly because it makes an acute angle with $\mu_2 = 1$) and the third quadrant solution is k_{2z}^L. Our prescription to establish k_{2z}^R for a true vacuum (i.e., $\epsilon_2'' = 0$) is to take the limit $\epsilon_2'' \to 0$, which yields k_{2z}^R as the solution along the positive imaginary axis.

Beware that if one adds a small amount of gain rather than loss to medium two, so that $\epsilon_2 = 1 + i\epsilon_2''$ where $\epsilon_2'' < 0$, then the two solutions for k_{2z} exist in the second and fourth quadrants as shown in Supp. Fig. 1(b), and in this case k_{2z}^R points predominantly along the *negative* imaginary axis. Thus, we see that the two limiting cases as gain or loss approaches zero do not yield the same result:

$$\lim_{\epsilon_2'' \to 0^+} k_{2z}^R = - \lim_{\epsilon_2'' \to 0^-} k_{2z}^R. \tag{3}$$

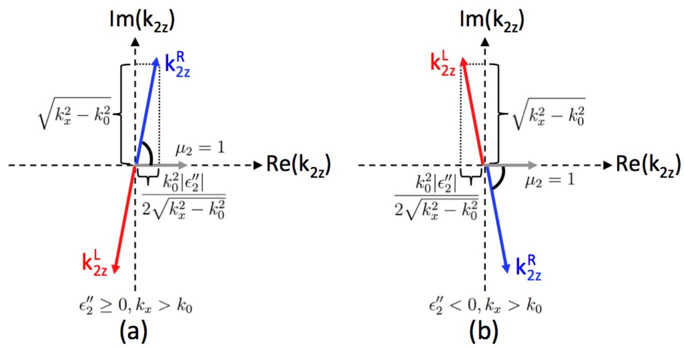

FIG. 1. Choosing the R or L label for an evanescent wave. (a) The two choices for k_{2z} are shown for the case of a slightly "lossy vacuum" ($\epsilon_2 = 1 + i\epsilon_2''$ where $\epsilon_2'' > 0$, $\mu_2 = 1$), for the case $k_x > k_0$. The first quadrant solution carries energy to the right and is labeled k_{2z}^R, and our prescription is to take the limit $\epsilon_2'' \to 0$ to determine that k_{2z}^R in the lossless case is along the positive imaginary axis. (b) For a slightly "gainy vacuum" ($\epsilon_2'' < 0$) the two solutions for k_{2z} are in the second and fourth quadrants, and k_{2z}^R approaches the negative imaginary axis as $\epsilon_2'' \to 0$. The magnitudes of the real and imaginary parts of k_{2z} in (a) and (b) are approximated using the first order Taylor expansion for small ϵ_2'': $k_0^2|\epsilon_2''| \ll k_x^2 - k_0^2$.

To have an unambiguous labeling convention for the case $\epsilon_2'' = 0$, we emphasize that one must take the limit as *loss* approaches zero, which can be different from the limit as gain approaches zero in the case of evanescent waves.

Finally, it is worth noting that this discontinuity in the two limiting cases, apart from being a footnote in establishing a labeling convention, is actually at the heart of the debate over single-surface amplified TIR. When medium two has gain, if one chooses k_{2z}^R as the transmitted wavevector (in accordance with our postulate), then it seems unphysical that as the gain approaches (but does not reach) zero the transmitted wave should still be strongly amplified. To remedy this situation it has been suggested that the correct choice for the transmitted wavevector should be k_{2z}^L when medium two has gain and $k_x > k_0$, so that the transmitted wave decays in the $+z$-direction. We believe, instead, that the discontinuity in the two limits is not as unphysical as it might appear at first: the transmitted wave propagates a large distance in the x-direction while barely moving forward in the z-direction (since $k_x \gg \mathrm{Re}(k_{2z}^R)$), so the large gain in the z-direction is actually a result of the long propagation distance along the x-direction. Far more unphysical, in our opinion, is the decision to switch the transmitted wavevector from k_{2z}^R when $k_x < k_0$ to k_{2z}^L when $k_x > k_0$.

All of these arguments aside, however, the purpose of our paper has been to demonstrate a mechanism by which the specularly reflected beam from a finite-thickness slab can be amplified, both below and above the critical angle.

II. GAIN SATURATION

Physicists familiar with the principles of lasers should be rightfully wary of a steady-state solution with roundtrip coefficient $|\nu|$ greater than one. When an active medium is pumped hard enough to generate a population inversion large enough to yield $|\nu|$ greater than one, light initially generated by spontaneous emission in the cavity will be amplified after each roundtrip. However, the field amplitude does not grow without bound–when the field is large enough, the upper state lifetime is reduced by stimulated emission which causes the population inversion to decrease to a level such that $\nu = 1$, resulting in a steady-state lasing solution. This phenomenon of gain reduction with increasing field amplitude is known as gain saturation. In a laser, therefore, the situation $|\nu| > 1$ is only a transient state. It's obvious that it cannot be a steady-state solution, because the field would grow without bound.

The situation changes when we allow an incident wave to strike the active medium, as we do in this Letter. Note that ν is defined as the roundtrip coefficient *in the absence of an incident wave*; that is, the reflectivity r_{21} is calculated by assuming that there is no wave in medium 1 arriving at the cavity. We can also define an effective reflectivity $r_{21}^{\text{eff}} \equiv E_2^R/E_2^L$ which takes into account the effect of the incident wave. Similarly, we could define an effective roundtrip coefficient in the slab which replaces r_{21} with r_{21}^{eff}: $\nu^{\text{eff}} = r_{21}^{\text{eff}} r_{23} \exp(2ik_{2z}^R d)$. We emphasize that *every possible steady-state solution to the problem under consideration, whether the slab is passive or active and whether there is an incident wave or not, satisfies the condition* $\nu^{\text{eff}} = 1$. This is a property of steady-state solutions: the field in the slab must regenerate itself after every roundtrip, taking into account all sources and sinks. Therefore, in solutions where $|\nu| > 1$, the incident wave must interfere destructively with the field in the slab so that $|r_{21}^{\text{eff}}| < |r_{21}|$ and ultimately force ν^{eff} to equal 1. In summary, when there is no incident wave the situation $|\nu| > 1$ is temporary because the field will grow until gain saturation (a nonlinear effect) forces the $\nu = 1$ solution. With an incident wave, a linear steady-state solution is possible even when $|\nu| > 1$ because of the reduction in the effective facet reflectivity r_{21}^{eff}, which prevents the unbounded growth of the fields so that we do not have to rely on gain saturation to avoid a nonphysical divergence.

III. PULSE OF LIGHT INCIDENT ON GAINY SLAB WITH $|\nu| > 1$

The video file pulse_video.avi included online is a time-lapse video of $|E_y|^2$ of a pulse of light, rather than a beam, for the same material parameters as in Fig. 2(b) of the main text: $\epsilon_1 = \epsilon_3 = 2.25$, $\epsilon_2 = 1 - 0.01i$, $\mu_1 = \mu_2 = \mu_3 = 1$, $d = 28$ μm. The white vertical lines in the video identify the 1-2 and 2-3 interfaces. The incident pulse is s-polarized and Gaussian in both space (FWHM = 13.3 μm) and time (FWHM = 50 fs, or 15 optical cycles). The central wavelength of the pulse is $\lambda_o = 1$ μm, and the mean incidence angle (i.e., averaged over all constituent planewaves) is 30°. The size of each video frame is 210 μm by 150 μm (height by width). The time elapsed between frames is 10 fs, and the entire video spans 1.22 ps (123 frames total). The field $|E_y|^2$ is plotted on a logarithmic scale covering 3 decades, i.e. red corresponds to the maximum intensity and blue corresponds to intensities less than or equal to 1/1000th of the maximum. The background in this image is blue, which corresponds to the minimum of $|E_y|^2$, whereas the background in Figs. 2(a) and 2(b) is green because it is the field E_y that was plotted in that case, so that blue corresponded to the maximum negative field.

In the video, one first sees the incident pulse near the bottom left of the screen, traveling up and to the right. The pre-excitation is soon seen in the slab at the bottom of the frame, and the reflected pulse that corresponds to the $m = 1$ term in the primed partial wave expansion leaves the slab and propagates up and to the left in medium one. The pre-excitation in the slab then undergoes one roundtrip as it zig-zags upward, giving rise to a transmitted pulse in medium three followed by the $m = 0$ reflected pulse in medium one. The pre-excitation then makes one more roundtrip, giving rise to another transmitted pulse in medium three, and then approaches the 2-1 interface at the same time the incident pulse arrives from the other side. The two pulses interfere in such a way as to yield an amplified specularly reflected pulse by entirely depleting the energy content of the slab. The fact that the pre-excitation in the slab travels in the $+x$-direction clearly distinguishes this behavior from negative refraction.

IV. DESCRIPTION OF SIMULATION

The E-field plots of the Gaussian beams and the video of the pulse were created using MATLAB. The field at each pixel is determined by superposing a large (but of course finite) number of planewave solutions. In this regard, the plots represent perfectly analytical solutions to Maxwell's equations.

As described in the main text, the response of the slab to an incident s-polarized planewave with amplitude E_1^R and wavevector $\boldsymbol{k}_1^R = k_x\hat{\boldsymbol{x}} + k_{1z}^R\hat{\boldsymbol{z}}$ is given by

$$E_y(x,z) = \begin{cases} E_1^R \exp(ik_x x + ik_{1z}^R z) + E_1^L \exp(ik_x x + ik_{1z}^L z) & : z \le 0 \\ E_2^R \exp(ik_x x + ik_{2z}^R z) + E_2^L \exp(ik_x x + ik_{2z}^L z) & : 0 \le z \le d \\ E_3^R \exp[ik_x x + ik_{3z}^R (z-d)] & : z \ge d \end{cases} \tag{4}$$

and the time-dependence factor $\exp(-i\omega t)$ is not explicitly written. The wavevector components k_{2z}^R and k_{3z}^R are determined by the dispersion relation

$$k_{\ell z}^R = \sqrt{(\omega/c)^2\mu_\ell\epsilon_\ell - k_x^2}, \tag{5}$$

where μ_ℓ and ϵ_ℓ are the relative magnetic permeability and electric permittivity constants of material ℓ, and the sign of the square root is chosen according to the prescription described in Supplementary Sec. 1. The four unknown wave amplitudes are found by satisfying Maxwell's boundary conditions to be

$$E_2^R = \frac{2k_{1z}^R(k_{3z}^R + k_{2z}^R)E_1^R}{(k_{2z}^R + k_{1z}^R)(k_{3z}^R + k_{2z}^R) + \exp(2ik_{2z}^R d)(k_{3z}^R - k_{2z}^R)(k_{2z}^R - k_{1z}^R)} \tag{6}$$

$$E_2^L = \frac{-2k_{1z}^R(k_{3z}^R - k_{2z}^R)E_1^R}{(k_{2z}^R - k_{1z}^R)(k_{3z}^R - k_{2z}^R) + \exp(-2ik_{2z}^R d)(k_{3z}^R + k_{2z}^R)(k_{2z}^R + k_{1z}^R)} \tag{7}$$

$$E_1^L = E_2^R + E_2^L - E_1^R \tag{8}$$

$$E_3^R = E_2^R \exp(ik_{2z}^R d) + E_2^L \exp(-ik_{2z}^R d). \tag{9}$$

To construct the Gaussian beam from the planewave solutions, we begin by expressing E_y in the $z = 0$ plane for a beam traveling parallel to the z-axis

$$E_y(x, z=0) = E_0 \exp\left(-\frac{x^2}{2\sigma_x^2}\right), \tag{10}$$

where E_0 is the peak amplitude and σ_x is directly proportional to the spatial FWHM

$$w_x = 2\sqrt{2\ln 2}\sigma_x. \tag{11}$$

By Fourier transforming and subsequently inverting the transform, the field can equivalently be written as an integral in k-space,

$$E_y(x, z = 0) = \int_{-\infty}^{\infty} dk_x E_1^R(k_x) \exp(ik_x x), \tag{12}$$

where

$$E_1^R(k_x) = \frac{E_0 \sigma_x}{\sqrt{2\pi}} \exp\left(\frac{-k_x^2}{2(1/\sigma_x)^2}\right), \tag{13}$$

and the FWHM in k-space is

$$w_k = 2\sqrt{2\ln 2}/\sigma_x. \tag{14}$$

To propagate the beam beyond the $z = 0$ plane, we associate with each value of k_x a component k_{1z}^R such that the total wavevector obeys the dispersion relation in medium 1,

$$k_{1z}^R(k_x) = \sqrt{(\omega/c)^2 \mu_1 \epsilon_1 - k_x^2}. \tag{15}$$

Now the Gaussian beam can be expressed as a function of x and z by

$$E_y(x, z) = \int_{-\infty}^{\infty} dk_x E_1^R(k_x) \exp[i(k_x x + k_{1z}^R z)]. \tag{16}$$

At this point, we must approximate the integral in Eq. 16 by discretization so that the calculation can be carried out by a computer. We restrict k_x to a finite sampling width given by $-w_s/2 < k_x < w_s/2$, and sample the beam equidistantly within this region with a total number of samples N_s. The integral in Eq. 16 is approximated by the sum

$$E_y(x, z) = \sum_{k_x = -w_s/2}^{w_s/2} \Delta k_x E_1^R(k_x) \exp[i(k_x x + k_{1z}^R z)], \tag{17}$$

where

$$\Delta k_x = \frac{w_s}{N_s - 1}. \tag{18}$$

At this point, it is helpful to think of E_1^R, k_x, and k_{1z}^R as vectors containing N_s numerical elements each. To rotate the beam so that it travels at an angle θ to the z-axis, we perform the transformation

$$k_x \to \cos(\theta) k_x + \sin(\theta) k_z \tag{19}$$

$$k_{1z}^R \to -\sin(\theta) k_x + \cos(\theta) k_{1z}^R \tag{20}$$

on each element of k_x and k_{1z}^R. (The Fourier amplitude of each plane-wave $E_1^R(k_x)$ is unaffected by the rotation in the case of s-polarized light.) Finally, to displace the waist of the beam to some location (x_0, z_0) in the incidence medium, one must multiply each Fourier amplitude by

$$E_1^R(k_x) \rightarrow E_1^R(k_x)\exp[-i(k_x x_0 + k_{1z}^R z_0)]. \tag{21}$$

With these redefined values for E_1^R, k_x, and k_{1z}^R, the sum in Eq. 17 is a good approximation to a Gaussian beam traveling at an angle θ whose waist is located at (x_0, y_0). The total E-field at any point in the system is given by

$$E_{\text{tot}}(x,z) = \begin{cases} \text{Real}\{\sum \Delta k_x \left(E_1^R(k_x)\exp[i(k_x x + k_{1z}^R z)] + E_1^L(k_x)\exp[i(k_x x + k_{1z}^L z)]\right)\}, & z \leq 0 \\ \text{Real}\{\sum \Delta k_x \left(E_2^R(k_x)\exp[i(k_x x + k_{2z}^R z)] + E_2^L(k_x)\exp[i(k_x x + k_{2z}^L z)]\right)\}, & 0 \leq z \leq d \\ \text{Real}\{\sum \Delta k_x E_3^R(k_x)\exp[i(k_x x + k_{3z}^R z)]\}, & z \geq d \end{cases} \tag{22}$$

where E_1^L, E_2^R, E_2^L, and E_3^R are calculated element-wise from $E_1^R(k_x)$ according to Eqs. 6-9. The beam plots in Fig. 2 of the main text are calculated pixel-by-pixel from the sum in Eq. 22, with the values of x and z indicating the location of the pixel. The resultant field is normalized to the maximum field value in the image, and displayed in color. The pulse video is calculated similarly, except that the field is Gaussian in space and time, and so the field must be sampled in both spatial and temporal frequency. The calculation time is significantly longer for the pulse compared to the beam, and the simulations are only practical on a supercomputer.

The finite nature of the sampling has consequences which must be considered in order to be sure that our results are not affected by numerical artifacts. Firstly, the truncation of the Gaussian beam in k-space to the sampling width w_s leads to a convolution with a sinc function in the spatial domain. Therefore, the side-tail of our beam is not truly Gaussian; rather, the envelope of the side-tail is Gaussian but the side-tail itself exhibits periodic sinc-like fluctuations in intensity (which cannot be seen in Fig. 2 of the main text, but can be seen in logarithmic plots which resolve the small intensities of the side-tail). The sampling width chosen for Fig. 2 was $w_s = 2w_k$ (with $N_s = 501$). We made sure that other choices of the sampling width, $w_s = 3w_k$ and $4w_k$ (with proportionally larger N_s so that Δk_x remained constant), did not affect the behavior of the plots. Therefore, our conclusions are not affected by the precise value of the sampling width w_s. Secondly, the finite number of samples N_s implies the spectrum of k_x values is discrete, so the incident beam is periodic in space. This means that in the plots of Fig. 2 in the main text, there is not just one incident beam but an infinite number of them impinging on the slab, spaced periodically along the x-axis by a distance $2\pi/\Delta k_x = 2830$ μm. If the sampling is increased from $N_s = 501$ to 2001 (while keeping $w_s = 2w_k$ constant), the distance between adjacent beams increases to 11330 μm, but the plots in both Figs. 2(a) and 2(b) look identical to the ones with 501 samples. Therefore, 501 samples is sufficient in this case to ensure the beams do not interfere with each other, and the plot is a good representation of the field of a single beam.

V. FUTURE WORK

Our intent in this Letter has been to demonstrate the unintuitive behavior of a beam incident on a slab whose roundtrip coefficient is greater than one, which we demonstrated using analytical solutions to Maxwell's equations. We argued that when $|\nu| > 1$, the amplification of typically negligible field amplitudes results in a "pre-excited" beam in the slab which interferes with the incident beam to prevent the divergence of the field. We focused on two peculiar consequences of this phenomenon: 1) the specularly reflected beam is amplified only when $|\nu| > 1$, and 2) the field in the slab is dominated by the wavevector k_{2z}^L when $|\nu| \gg 1$. Our analysis has been restricted to the case of planar media with infinite extent in the x and y-dimensions, with homogeneous and frequency-independent material parameters. We do not consider this a serious drawback of our argument, as the majority of analyses of the three-layer problem employ the same assumptions. From a purely theoretical viewpoint and within the confines of the assumptions we have made, therefore, we believe that these results provide perspective for single-surface amplified TIR, and counter the notion of negative refraction in a nonmagnetic slab. Because the pre-excitation mechanism can only occur in a finite-thickness slab, we speculate that k_{2z}^R is the correct choice for the transmitted wavevector in the single-surface problem in all cases; in other words, the transmitted wave always carries energy away from the interface. There remain many open questions to be answered theoretically. Finite-difference time-domain simulations may be best suited to determine how the pre-excitation mechanism, which begins at $x \ll 0$, is affected when the slab has a finite length in the x-direction. Spontaneous emission and gain saturation must be accounted for in real materials. How would the slab behave if the beam had a truly finite width and no side-tail? To investigate how the slab reaches the steady state over time in response to a pulse with a sharp turn-on, the material parameters of the slab must be made to obey the Kramers-Kronig relations. Despite our lack of answers to these important questions, we hope that our analysis has clarified at least some of the relevant issues.

Surprisingly, $|\nu|$ can exceed one even in passive media provided $|r_{21}|$ or $|r_{23}|$ exceeds one, which can happen for incident evanescent waves near surface plasmon resonances. Understanding the role of ν in these cases can yield additional insight, particularly to the case of Pendry's lens [1, 2], and will be treated in future work.

Second Report of Referee A, received July 16th, 2013:

In the revised manuscript the authors clarified some of the concerns that were raised in my original report. Having the details of the numerical calculations performed by the authors we were able to reproduce Fig. 2 in the manuscript. In fact, it appears that Fig. 2 and the movie presented in this work are not the artifact of Fourier series.

However, having read the reply, the accompanying supplementary information, and having experimented with the system under consideration, I still believe that the solution that the authors present is not physical one.

The authors do rely on plane wave solutions of Maxwell equations. Plane waves, by their nature, spread throughout time and space; therefore, any solution derived from the plane wave spectrum will finite in time and space. The system with net roundtrip gain is a lasing laser that does not have a steady state solution and should not be represented as plane wave spectrum. In reality, the system will be "pre-excited" with spontaneously emitted photon that will immediately reduce the gain level |nu| to 1 or even further.

Nevertheless, the solution shown in the work does represent a solution to some problem. The question is what that problem is. It appears to me that this problem is the problem of reflection of the beam from LOSSY material, with imaginary part of permittivity given by e"=-0.01i, with plane wave convention written as exp(+iwt-ik.r).

The experiment that we did is increasing the "gain" parameter, given by imaginary part of the permittivity. Interestingly, as we increase the gain, we can clearly see that "transmitted" beam disappears, while amplitude of the "specular reflected" beam converges to some value (greater than 1). This fact seems to contradict the pre-excitation hypothesis. Moreover, this fact seems to contradict the increase of the gain itself. However, the results of these calculations are identical to the results of calculating the pulse reflection from LOSSY material, if one "mislabels" incident and reflected beams.

Therefore, I believe that the solution shown in the movie corresponds to the physical situation where the planar, lossy, slab is excited by the beams incident from the top of the screen, and propagating downward; the time axis in the calculation (or, equivalently, frequency) has to be inverted.

First Report of Referee B, received July 16th, 2013:

The authors present a theoretical study of Fresnel reflection from a cavity with net roundtrip gain. I have read the paper and the comments of the referees, and the replies by the authors. The work is certainly interesting and controversial. A major concern appears to be the numerical verification via a full time-domain solution of Maxwell's equations, such as FDTD; here I am puzzled by the fact that a good PML does not allow the authors to study the problem at hand. From personal experience, this usually means that the predictions are too sensitive to see, and can be washed out or may not be real.

Additionally, it is well known that the medium must satisfy causality for these regimes (for example a negative index cannot be simulated in FDTD unless a causal model is used), yet the chosen eps does not include material dispersion. Before recommending this paper for publication, I would like to ask for a full section on FDTD simulations of this prediction and some repeated calculation with a dispersive medium (within the current technique and with the FDTD simulations) – i.e., one that satisfies causality and the KK relations. These realistic details are necessary to have confidence in these results and can largely be put in the Supplementary Information.

Response of authors, sent August 23^{rd}, 2013:

Dear Dr. Malenfant and Referees:

Thank you for your consideration of our manuscript, as well as the critical analyses by the referees. We have revised the manuscript and its supplemental information as listed briefly below and described in detail later.

1) We emphasize in the main text that we are only presenting the Fresnel solution for the given geometry. In the concluding paragraph, we mention that there can certainly exist other solutions to the full time-dependent Maxwell's equations, some of which may diverge. We have included a reference to a work that explores this divergent solution using FDTD methods [4], but we have also included references to experiments [14, 20, 21] that have probed the regime $|\nu|>1$ and have not suffered from instabilities.
2) While reviewing the literature, we noticed that the mathematical duality between the roundtrip coefficient ν and its inverse that we have presented in the paper provides a simple explanation for the localization of light in both lossy and gainy media, an effect discovered by others [4-8] and now referenced in our introductory section.
3) We have added Fig. S3 to the Supplementary Information to better explain that our Fig. 2(b) is not going to change by finer sampling in the Fourier domain.
4) We have added a discussion to the Gain Saturation section of the Supplementary Information on a laser medium with self-feedback. This is an experimentally proven situation in which an amplifying cavity with roundtrip coefficient greater than 1 is prevented from lasing by the influence of a field incident on the cavity.

Before we explain the above revisions in detail in conjunction with our specific responses to each of the referees' comments, we would like to make a general statement about the intent of our manuscript. We hope that by explaining our approach and the questions that we have set out to answer (and those we have not set out to answer), the referees will be in a better position to judge the value of the work. Certainly, we have not done a good job of explaining these points so far, both in the original manuscript and in our previous letter to the first referee. Hopefully this letter, together with the revisions to the manuscript, can fix these shortcomings.

We have intentionally titled our paper "Fresnel reflection from a cavity with net roundtrip gain" as opposed to simply "Reflection from a cavity with net roundtrip gain." The Fresnel formalism is a specific prescription for understanding reflection and transmission of light, a formalism that relies exclusively on planewave solutions of Maxwell's equations—solutions that we all agree are unphysical due to their infinite extent in both space and time—yet have nevertheless been used by physicists for over a century to successfully explain experimental results. It turns out that this formalism admits a convergent solution for cavities with net roundtrip gain. Considering the monumental successes of the Fresnel formalism in treating other cases, it seemed reasonable to apply the *unvarnished* Fresnel formalism to this problem and investigate its predictions. The dearth of publications that take the formalism seriously, and apply it without introducing unwarranted assumptions, seems to have stemmed from the prevailing belief, seemingly supported by the geometric-series method of deriving the reflection coefficient, that the reflection coefficient must diverge when the cavity has net roundtrip gain. However, we did find in the literature many instances in which a cavity with $|\nu| > 1$ did play a role, although often this role went unnoticed. We mention these cases in the introduction of our manuscript, which include Pendry's lens, negative refraction in gainy non-magnetic media, and total internal reflection from an amplifying slab. (While revising the manuscript this second time, we also discovered another body of work on the duality between lossy and gainy media, both of which can lead to the localization of light, that is also related to a cavity with $|\nu| > 1$.) Due to this lack of awareness and appreciation for what can happen when the roundtrip coefficient exceeds one, our intent in this manuscript is simply to demonstrate the Fresnel solution for a cavity with $|\nu| > 1$. Of course, the Fresnel reflection coefficient for a planewave is easy to calculate, but we believe that we have lent valuable insight to the problem by considering the case of a finite-diameter beam incident obliquely on the cavity in Fig. 2(b). Furthermore, we have shown that this counterintuitive result is in fact consistent with the sum-over-partial-waves method; once the exchange $k_{2z}{}^R \rightarrow k_{2z}{}^L$ is made in Eq. 6, the resulting "primed" partial wave expansion in Eq. 7 provides a beautiful mathematical description of the plot in Fig. 2(b).

These results alone shed light on the four topics mentioned in the introduction. We have shown that negative refraction is *not* responsible for the peculiar location of the transmitted and reflected beams, as others have claimed. We have also shown that amplified total internal reflection from a slab falls within the regime $|\nu| >> 1$, and that the accompanying pre-excitation mechanism in this regime can explain the amplification of the specularly reflected beam. In this revised version of the manuscript, we also point out that the duality we have demonstrated between ν and $1/\nu$ in the partial wave expansion is a simple mathematical explanation for the duality between lossy and gainy media observed by others. Finally, the situation $|\nu| > 1$ can even occur in passive media for incident evanescent waves which excite surface plasmon resonances; the analysis we have presented in this paper therefore offers insight into the behavior of Pendry's lens, but there is not enough space to include these results in the present manuscript.

These three results, in particular the ones on negative refraction and amplified total internal reflection, are the main conclusions of the paper. We feel that it is important to point out that neither of the referees has so far questioned these conclusions (aside from the duality between loss and gain, which they will now be seeing for the first time). It is by no means a shortcoming of our analysis that it is based on planewave decomposition, on homogeneous media with infinite extent in the x and y directions, and on modeling gain with a refractive index whose imaginary part is negative. *These are fundamental tenets of the successful Fresnel formalism*, and furthermore, these are the exact same techniques that have been used over and over in the literature to address the problems of negative refraction in gainy media, amplified TIR, loss/gain duality, and Pendry's lens. All that we have done here is explore the logical consequences of the fundamental assumptions upon which such analyses have been based. We are not inventing new physics here, nor are we deploying new mathematical tools and techniques. We simply start with the assumptions underlying the Fresnel formalism, explore a range of parameters such as thickness and incidence angle, and describe the inexorable logical conclusion of these methods. This is the main contribution of our manuscript.

If the referees do not consider the above conclusions worthy of publication in Physical Review Letters, we will accept their judgment. We would like, however, to appeal to their sense of fairness in judging scientific works. We ask them to kindly consider the soundness of our theoretical arguments, which is the core of the paper, and the validity of our numerical simulations, which one of the referees has already managed to reproduce. Is this an important problem? Does our analysis build on previous work on the same topic? Does our analysis reveal something new that could potentially be worthy of further exploration? Have we made mistakes in our theoretical and/or numerical analyses? Will the subset of the physics community that cares about such problems find something worthwhile in this work? These are the questions that we hope the referees would ask themselves in evaluating our work.

The issues that the referees have raised so far, while important and worth pursuing in the future, are ancillary to our purpose with this manuscript, namely, to demonstrate the predictions of the Fresnel formalism for a cavity with net roundtrip gain. The first referee, by bringing up the effect of spontaneous emission, has questioned whether such a cavity can even exist. We believe that such a cavity can exist, as demonstrated by experiments done on the amplification of evanescent waves. We cite three such experiments in our paper, all of which fall in the $|\nu|>1$ regime, and none of which were affected by instabilities caused by spontaneous emission. We leave it to future experiments to determine whether solutions of the type presented in Fig. 2(b) can be seen in the laboratory. If they can, it would be an interesting effect and hopefully applications will follow. If they cannot, that simply means that the convergent Fresnel solution is not realizable, or is masked by an unstable solution. But this does not detract from our exposition of the Fresnel solution, a solution that is certainly of theoretical interest considering the widespread success of the Fresnel formalism in other regimes. Similarly, the second referee has asked us to confirm our numerical results by using the FDTD method. While we too would be curious to know whether the FDTD method will yield the same result as the Fresnel solution, we simply do not have sufficient expertise in the FDTD method to navigate the intricacies of commercial software, especially when the problem promises to be sensitive to numerical dispersion and noise arising from discretization, which would require careful tuning of the simulation parameters. Scientific progress is incremental, and we hope that our presentation of the Fresnel solution will allow someone with more expertise than us in the FDTD method to investigate the problem further.

In summary, since the referees seem to agree that our analysis has raised new questions worth pursuing, we ask that they judge our work based not on the questions we have left unanswered, but rather on the questions we have decisively answered by exploring in detail the Fresnel solution for a cavity with net roundtrip gain. If they don't approve of the manuscript, we will respect their decision to reject. However, if they see enough value in bringing this admittedly narrow but logical consequence of Maxwell's equations to the attention of the community, then we will be extremely grateful for not letting the fruits of our labor to go to waste.

We now proceed to respond specifically to the comments and suggestions made by the referees. We have included pdf documents of the main text and supplementary information in which new material included in this revision appears highlighted.

Referee 1:

In the revised manuscript the authors clarified some of the concerns that were raised in my original report. Having the details of the numerical calculations performed by the authors we were able to reproduce Fig. 2 in the manuscript. In fact, it appears that Fig. 2 and the movie presented in this work are not the artifact of Fourier series.

Thank you. We very much appreciate the time and effort that you have invested in verifying our numerical simulations, and hope that this has given you greater confidence not only in our numerical results, but also in the analytic Fresnel solution on which the numerical results are based.

However, having read the reply, the accompanying supplementary information, and having experimented with the system under consideration, I still believe that the solution that the authors present is not physical one (sic). The authors do rely on plane wave solutions of Maxwell equations. Plane waves, by their nature, spread throughout time and space; therefore, any solution derived from the plane wave spectrum will finite (sic) in time and space. The system with net roundtrip gain is a lasing laser that does not have a steady state solution and should not be represented as plane wave spectrum.

You may be right; the result is physical only if it can be demonstrated experimentally. We are presenting these results not as established facts, but as logical consequences of Maxwell's equations in conjunction with assumptions that we have clearly laid out concerning the nature of the gain medium and the properties of the incident beam. We hope that you'd agree, however, that the same methods give the correct (physical) result whenever $|\nu|<1$, as shown in Fig. 2(a). We also hope you'd agree that there is some value and interest in understanding what the Fresnel formalism predicts when $|\nu|>1$, even if the result turns out to be masked by an unstable solution, something that will hopefully be answered by future experiments.

However, we feel that it is unfair to assert that any solution derived from planewaves will be infinite in time and space. This statement is essentially dismissing all of Fourier transform theory, and we don't believe the physics community will cast us out as heretics if we put our foot down and say that Fourier transform theory is absolutely capable of creating functions that are finite in time and/or space. Granted, our plot in Fig. 2(b) is the result of a Fourier series, not a Fourier transform, as you pointed out in your first referee report. We argued before that the particular choice of sampling rate in the Fourier domain does not affect the appearance of Fig. 2(b), provided it is large enough, and now we have elaborated this point with an extra plot in the Supplementary Information, Fig. S3. Perhaps you have reproduced this result yourself. We are confident that you could make the sampling rate as large as you like, and Fig. 2(b) would look the same; by extension, we are confident that an analytical calculation of the Fourier integral, if it could be done, would yield the same field pattern as our Fourier series 'approximation.' The reason we are so confident—besides the fact that we have played with the sampling rate and confirmed these results—is that the theory we have laid out in Eq. 7 perfectly predicts the observed behavior of the beam. We respectfully urge both referees to please pay particular attention to the predictive power of this formula as discussed in the manuscript; we consider this to be an important result, and so far neither of the referees has commented on it.

With regard to the statement "the system with net roundtrip gain is a lasing laser," we do not think this is as self-evident as the referee believes. We have previously explained in the Gain Saturation section of the Supplementary Information that the incident wave, when transmitted into the cavity, destructively interferes with the field circulating in the cavity to prevent the field from diverging. When one accounts for this incident wave, the effective roundtrip coefficient in the slab, ν^{eff}, is exactly equal to one. To better explain this point, we have now added to the Gain Saturation section of the manuscript an example of an experimentally proven situation in which this exact effect takes place, which we hope the referees will appreciate. We consider a laser cavity defined by a slab of material and the facet mirrors, as in a semiconductor laser. An external mirror is placed next to one of the facets, thus leaving an air gap between one of the laser facets and the mirror. The light that leaves the laser is reflected by the mirror and returns to the laser. For certain values of the external mirror reflectivity and distance from the laser facet, this "self-feedback" light interferes destructively with the field circulating in the laser cavity. This reduces the "effective" facet reflectivity of the laser. Thus, it is possible for the laser to be pumped to a level such that the gain medium has $|\nu|>1$, yet the laser will not lase due to the destructive interference from the self-feedback. We hope the referees appreciate the analogy between this situation and the situation we have presented in the article: when light is incident externally on a cavity, it is ν^{eff}, not ν, that determines whether the cavity will lase.

In reality, the system will be "pre-excited" with spontaneously emitted photon that will immediately reduce the gain level |nu| to 1 or even further.

There is definitely not a strong enough consensus in the community on this subject for the referee to assert what will happen in reality, although we certainly understand why the referee's intuition led him to this conclusion. However, there already exists some experimental evidence that spontaneously-emitted photons may not destabilize the system. The experiments that have probed the regime $|\nu|>1$ [14,20,21] have *not* suffered from any instabilities caused by spontaneous emission or otherwise. We look forward to more experiments in the future to probe a broader range of parameters. We have added a paragraph to the Gain Saturation section in the Supplementary Information to explain the difference between photons propagating along the z-axis and those traveling obliquely to the z-axis. There is clearly a cavity for optical feedback for photons propagating along z; however, as explained in the manuscript, the roundtrip coefficient for these photons is *less than* one. For the parameters used in Fig. 2(b), only photons associated with incident angles greater than 27.34° will have $|\nu|>1$. Thus, a photon spontaneously emitted along z will *not* cause lasing. For those spontaneous photons emitted obliquely such that $|\nu| > 1$, the finite size of the slab in the x-direction (as it must be in an experiment) will allow the photons to leave the system. For a more detailed explanation, please see the paragraph at the end of the Gain Saturation section in the Supplementary Information.

Nevertheless, the solution shown in the work does represent a solution to some problem. The question is what that problem is. It appears to me that this problem is the problem of reflection of the beam from LOSSY material, with imaginary part of permittivity given by e”= 0.01i, with plane wave convention written as exp(+iwt ik.r). The experiment that we did is increasing the “gain” parameter, given by imaginary part of the permittivity. Interestingly, as we increase the gain, we can clearly see that “transmitted” beam disappears, while amplitude of the “specular reflected” beam converges to some value (greater than 1). This fact seems to contradict the pre-excitation hypothesis. Moreover, this fact seems to contradict the increase of the gain itself. However, the results of these calculations are identical to the results of calculating the pulse reflection from LOSSY material, if one “mislabels” incident and reflected beams. Therefore, I believe that the solution shown in the movie corresponds to the physical situation where the planar, lossy, slab is excited by the beams incident from the top of the screen, and propagating downward; the time axis in the calculation (or, equivalently, frequency) has to be inverted.

Yes, your numerical findings are exactly right. As you increase the gain, ν gets larger, and $1/\nu$ correspondingly gets smaller. Since you are probing a regime where $|\nu| > 1$, you can use the primed partial wave expansion in Eq. 7 to interpret your results. (Everything we have said in the manuscript with regard to increasing the thickness d of the gain layer applies also to the case of increasing the gain parameter, since both changes result in an increase in $|\nu|$. Please read the second paragraph on the fourth page of the manuscript for a more detailed explanation.) As you increase the gain, $1/\nu$ approaches zero, and so the contribution of each term in the geometric series becomes less apparent. This is why you only see the specularly reflected beam, which is governed by the term r'_{21}—the one term that is not part of the geometric series. However, to say that the terms of the geometric series become less apparent is vastly different from saying that they do not exist. The terms are non-zero, so if you were to look at the field pattern in the slab with a magnified scale (i.e., on a log plot), you would definitely still see the field zig-zagging up the slab from $x << 0$ up to $x = 0$. Therefore, this does not contradict the pre-excitation hypothesis. It also does not violate the concept of gain itself; it only runs against your intuitive notion of gain. Please reconsider your intuition that “more gain implies more energy.” The cavity in fact emits the most energy when ν is close to one—either above or below one—due to the resonant denominator in the Fresnel reflection (and transmission) coefficient. When ν is much smaller or much larger than one, the cavity is off-resonance; this is the essence of the loss/gain duality observed by others, and that we have now pointed out in our manuscript as well. Since we are used to dealing with situations in which $|\nu|$ is less than or equal to one, our intuition that “more gain implies more energy” is almost always correct. However, here we are dealing with a situation that is unlike most of the situations from which we have drawn our intuition; therefore, we should be willing to reevaluate that intuition. The primed partial wave expansion in Eq. 7 teaches us that when $|\nu| > 1$, the larger $|\nu|$ becomes, so too becomes larger the disparity in the magnitudes of successive terms in the partial wave expansion. Please understand that we would be the first to admit that the plot in Fig. 2(b) is counterintuitive! We are simply trying to explain different aspects of the solution to the best of our ability.

Your suggestion that the time-axis should be inverted to generate a physical solution is quite insightful. Below we would like to demonstrate the soundness of this insight by examining its mathematical basis. We will then argue that your time-reversed solution and our original solution are equally valid solutions to the Fresnel problem. Consider Maxwell's equations in homogeneous media in the absence of free charge and current:

$$\boldsymbol{\nabla}\times \boldsymbol{E} = -\mu_o\mu(\partial \boldsymbol{H}/\partial t),$$

$$\boldsymbol{\nabla}\times \boldsymbol{H} = \varepsilon_o\varepsilon(\partial \boldsymbol{E}/\partial t),$$

$$\boldsymbol{\nabla}\cdot \boldsymbol{E} = 0,$$

$$\boldsymbol{\nabla}\cdot \boldsymbol{H} = 0.$$

Here ε and μ are the relative electric permittivity and magnetic permeability of the media, respectively. It is readily seen that the equations are unchanged if one performs the two transformations $\{t\rightarrow -t,\ \boldsymbol{H}\rightarrow -\boldsymbol{H}\}$ on Maxwell's equations. Now, suppose that, in addition to providing a time-lapse video of the field component E_y of the pulse, we had also provided videos of H_x and H_z (which we could have done using essentially the same code). Anyone who watched these three videos in reverse, while also changing the sign of H_x and H_z, would say that the videos obey Maxwell's equations. What is happening here? When we make the transformation $t\rightarrow -t$ but keep $\varepsilon=1-0.01\text{i}$, the slab actually becomes a lossy medium even though we have not changed its dielectric constant. The reason for this, as you have correctly explained, is that each planewave is now described by $\exp(\text{i}\boldsymbol{k}\cdot\boldsymbol{r}+\text{i}\omega t)$ rather than $\exp(\text{i}\boldsymbol{k}\cdot\boldsymbol{r}-\text{i}\omega t)$, and with this convention a negative imaginary part of the dielectric constant is associated with *lossy* media.

For instance, running our pulse video backwards would show multiple pulses arriving at different times at the 1-2 interface. First the specular pulse would arrive, then the $m=0$ pulse, then later the $m=1$ pulse. (There would also be many pulses arriving at the 3-2 interface; namely, the time-reversed transmitted pulses.) While this video is a valid solution to Maxwell's equations, it would be difficult to set up the corresponding system in practice, as it would require a sophisticated external source to control precisely the arrival times and positions of an infinite number of pulses at the 1-2 and 3-2 interfaces.

As you mentioned, you focused in your own simulations on the situation of very large $|\nu|$, for which the terms in the geometric series are very small, but again, not negligible. Therefore, we agree that the time-reversed video you describe would certainly *appear* to resemble that of an incident beam traveling down and to the right, impinging on a lossy slab. However, because the terms of the geometric series are non-zero, your solution would also require a sophisticated source to generate an infinite number of incident beams. Basically, just because the terms of the geometric series are small does not mean they are negligible. In order to properly apply the time-reversal argument, all of the terms of the geometric series—and therefore all of the reflected and transmitted pulses—need to be time-reversed. In summary, both your time-reversed solution (with *all* the reflected and transmitted beams reversed) and our solution are perfectly valid Fresnel solutions for the geometry and material parameters considered. Our solution corresponds to the case of a single incident beam arriving at the 1-2 interface, whereas your solution requires multiple beams arriving at the 1-2 and 3-2 interfaces.

Referee 2:

The authors present a theoretical study of Fresnel reflection from a cavity with net roundtrip gain. I have read the paper and the comments of the referees, and the replies by the authors. The work is certainly interesting and controversial.

Thank you. We are happy that you find the topic interesting.

A major concern appears to be the numerical verification via a full time-domain solution of Maxwell's equations, such as FDTD; here I am puzzled by the fact that a good PML does not allow the authors to study the problem at hand. From personal experience, this usually means that the predictions are too sensitive to see, and can be washed out or may not be real.

As stated earlier, our purpose in this manuscript is to present the Fresnel solution to the cavity with net roundtrip gain. Because the problem is analytically tractable (and you have not raised any issues with our theoretical analysis), we hope that you agree that we have correctly presented the Fresnel solution, which is the core of our argument. Therefore, we argue that our solution does not require numerical verification to prove that it is indeed a solution of Maxwell's equations.

We are by no means dismissing the value of FDTD simulations for other purposes; we just think it is not necessary to prove that the solution we have presented is the correct Fresnel solution. We realize that the Fresnel approach to solving for reflection and transmission coefficients implicitly begins with the time-harmonic subset of Maxwell's equations. Therefore, the Fresnel formalism can only find convergent solutions, whereas there may very well exist divergent solutions to the full time-dependent Maxwell equations. We have added a paragraph at the very end of the manuscript to emphasize this point.

For example, if you take the field distribution shown in Fig. 2(b) as an initial condition (and not worry about how that field got there in the first place), then evolve that field forward in time, we are quite confident that the *E*-field at each point in space will simply vary harmonically as $\exp(-i\omega t)$. However, it is another question entirely as to how one might create this field distribution in the first place, starting from a state with no electromagnetic fields anywhere in space. We believe FDTD solutions will be useful to help answer this latter question, by looking at the time evolution of a pulse with a sharp turn-on time incident on a slab with finite length in the *x*-direction. We have added a reference to a paper [4] that performed finite element simulations of a planewave with a sharp turn-on time incident normally on a slab with $|\nu| > 1$; they found that the E-field emitted by the slab grew exponentially with time, never converging. We did not find any references to FDTD simulations of pulses at an oblique incidence angle. We believe that such simulations could potentially show the pre-excitation mechanism, resulting in the transmitted beam originating from $x < 0$, even if the fields eventually diverge when the simulation runs for long enough time.

However, we are not experts in FDTD simulation techniques. To properly do these simulations would require either a personally developed FDTD code or a nuanced understanding of the commercial codes, neither of which we possess. We tried some preliminary FDTD simulations using a commercial code, but the problems that we observed with the imperfect PML affected us. Again, we reiterate that an FDTD demonstration of the solution we have presented is ancillary to the purpose of this manuscript. In the future, we will invest more time in FDTD simulations of this phenomenon to satisfy our own curiosity. Perhaps others more knowledgeable in the FDTD method will be inspired by our Fresnel solution to do their own investigation. After all, scientific ideas build on each other incrementally in most cases, and we can only hope that you will find our argument (which is essentially a theoretical argument based on the planewave expansion method and the convergence or divergence of the partial wave sum) worth publishing even without additional FDTD verification.

Additionally, it is well known that the medium must satisfy causality for these regimes (for example a negative index cannot be simulated in FDTD unless a causal model is used), yet the chosen eps does not include material dispersion. Before recommending this paper for publication, I would like to ask for a full section on FDTD simulations of this prediction and some repeated calculation with a dispersive medium (within the current technique and with the FDTD simulations) – i.e., one that satisfies causality and the KK relations. These realistic details are necessary to have confidence in these results and can largely be put in the Supplementary Information.

Figure 2(b) in our paper is the field distribution for a monochromatic beam of wavelength $1\,\mu\text{m}$. Because the field only contains one frequency, there is no need to account for dispersion. In other words, we have presented the correct Fresnel solution for a particular value of ε_2, which is the dielectric constant of medium two evaluated at the wavelength of $1\,\mu\text{m}$. We assume your request for an ε_2 that obeys the Kramers-Kronig (K-K) relations refers to our pulse simulation. We agree that the K-K relations are important for causality. However, the purpose of our pulse video is not to demonstrate the causal creation of the pre-excitation; such a demonstration would be impossible even if the slab dispersion did obey the K-K relations, since our simulations are not capable of creating a pulse with a sharp turn-on time. (Instead, as you mention, FDTD simulations using a slab that obeys the K-K relations is best suited for this investigation.) The purpose of our pulse video is simply to demonstrate that the pulse in the slab is zig-zagging in the $+x$ direction, to distinguish the behavior of the pulse from negative refraction. One could just as well look at the direction of the Poynting vector shown in Fig. 2(b) to grasp this point, but we included the pulse video because it reiterates the point. Finally, we emphasize once again that the theory in our paper very clearly shows that the "pre-excitation" behavior occurs for planewaves for which $|\nu| > 1$. Once you choose a dispersion function that satisfies the K-K relations, you can still identify the planewaves that satisfy $|\nu| > 1$, and it is the superposition of these planewaves that creates the "pre-excitation" behavior within the Fresnel formalism.

Finally, we would like to comment on an important difference between negative-index media (mentioned by the referee) and the slab we have studied. In the case of negative-index media, and also in cases where the refractive index is positive but less than one, it is imperative to incorporate the frequency dependence of $\varepsilon(\omega)$ into the simulations. This is simply because the phase velocity is either in the opposite direction to the direction of energy flow, or because it is greater than the speed c of light in vacuum. In both cases, the frequency-dependence of $\varepsilon(\quad)$ must be taken into account, so that the group velocity will show the correct physical behavior for realistic beams. In our case, however, the refractive indices are greater than or equal to unity, which means that one can always choose a sufficiently broad pulse (i.e., narrow frequency spectrum) and allow $\varepsilon(\quad)$ within that narrow range of frequencies to remain more or less constant. In other words, the group velocity under such circumstances could be made very nearly equal to the phase velocity, which is less than or equal to c and also in the same direction as that of the energy flow. Under such circumstances, the incorporation of the K-K relations into $\varepsilon(\quad)$ does not modify the qualitative behavior of the simulations.

Thank you again for your continued consideration of our manuscript.

Sincerely,

Tobias S. Mansuripur, Graduate Student
Department of Physics
Harvard University

Version 3 of the manuscript (submitted to PRL 8/23/2013).
Changes relative to version 2 are highlighted.

Fresnel reflection from a cavity with net roundtrip gain

Tobias S. Mansuripur[1,*] and Masud Mansuripur[2]

[1]Department of Physics, Harvard University, Cambridge, MA 02139
[2]College of Optical Sciences, The University of Arizona, Tucson, AZ 85721

A planewave incident on an active etalon with net roundtrip gain may be expected to diverge in field amplitude, yet applying the Fresnel formalism to Maxwell's equations admits a convergent solution. We describe this solution mathematically, and provide additional insight by demonstrating the response of such a cavity to an incident beam of light. Cavities with net roundtrip gain have often been overlooked in the literature, and a clear understanding of their behavior yields insight to amplified total internal reflection, negative refraction in nonmagnetic media, and negative-index lenses.

The Fresnel coefficients govern the reflection and transmission of light for the simplest possible scenarios: at planar interfaces between homogeneous media. Despite this simplicity, some interesting solutions have been discovered only recently, such as 1) the amplification of evanescent waves in a passive, negative-index slab [1–3] and 2) a duality between loss and gain leading to the localization of light in both cases [4–8]. In addition, controversy regarding the proper choice of the wavevector in active media has persisted in relation to the possibility of 3) negative refraction in nonmagnetic media [9–13] as well as 4) single-surface amplified total internal reflection (TIR) [14–19]. It turns out that all four of these cases share a common thread: the presence of a cavity whose roundtrip gain exceeds the loss. In this Letter, we explore more generally the response of such a cavity to an incident beam of light.

To begin, we establish a convention that allows us to more clearly discuss the direction of energy flow. We associate with each wavevector a superscript R (L) to indicate that the wave carries energy to the right (left), namely, that for which the time-averaged z-component of the Poynting vector is positive (negative). (See supplementary information for details.) For the single-surface problem, shown in Fig. 1(a), the incident wavevector in medium one is $\boldsymbol{k}_1^R = k_x\hat{\boldsymbol{x}} + k_{1z}^R\hat{\boldsymbol{z}}$, and the reflected wavevector is $\boldsymbol{k}_1^L = k_x\hat{\boldsymbol{x}} + k_{1z}^L\hat{\boldsymbol{z}}$, where $k_{1z}^L = -k_{1z}^R$. The real-valued component k_x, once determined by the incident wave, is the same for all wavevectors in the system. For the transmitted wavevector, the dispersion relation offers two choices for k_{2z},

$$k_{2z} = \pm\sqrt{(\omega/c)^2\mu_2\epsilon_2 - k_x^2}, \quad (1)$$

where ω is the angular frequency, c is the speed of light in vacuum, and μ_2 and ϵ_2 are the relative permeability and permittivity. It is universally agreed that the correct choice for k_{2z} in the single-surface problem is k_{2z}^R (i.e., that the transmitted energy flows away from the interface), irrespective of the material parameters or the nature of the incident wave, except possibly in the case of amplified TIR, for which there remains debate. Due to this controversy, let us postulate for now that k_{2z}^R is the correct choice in all cases, so that we can unambiguously define the single-surface Fresnel reflection and transmission coefficients

$$r_{\ell m} = \frac{\tilde{k}_{\ell z}^R - \tilde{k}_{mz}^R}{\tilde{k}_{\ell z}^R + \tilde{k}_{mz}^R}, \; t_{\ell m} = \frac{2\tilde{k}_{\ell z}^R}{\tilde{k}_{\ell z}^R + \tilde{k}_{mz}^R} \quad (2)$$

where we have generalized the result for incidence medium ℓ and transmission medium m. For s-polarization we have defined $\tilde{k}_{nz} \equiv k_{nz}/\mu_n$, while for p-polarization $\tilde{k}_{nz} \equiv k_{nz}/\epsilon_n$.

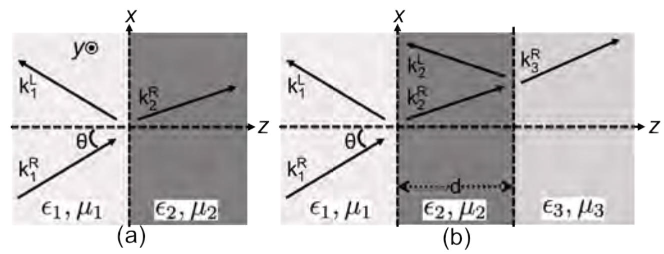

FIG. 1. Geometry of the (a) single-surface and (b) cavity problems. All media are infinite in the x and y-directions. The arrows denote the wavevectors of the planewaves present in each layer.

We now consider the case of light incident on a cavity, shown in Fig. 1(b). The total E-field resulting from an s-polarized incident wave in medium one with amplitude E_1^R is given by

$$E_y(x,z) = \begin{cases} E_1^R \exp(ik_xx + ik_{1z}^R z) \\ \quad +E_1^L \exp(ik_xx + ik_{1z}^L z) & : z \le 0 \\ E_2^R \exp(ik_xx + ik_{2z}^R z) \\ \quad +E_2^L \exp(ik_xx + ik_{2z}^L z) & : 0 \le z \le d \\ E_3^R \exp[ik_xx + ik_{3z}^R (z-d)] & : z \ge d \end{cases} \quad (3)$$

where the time-dependence factor $\exp(-i\omega t)$ has been omitted. The most direct route to solve for the four unknown wave amplitudes is to enforce Maxwell's boundary conditions at $z = 0$ and $z = d$, which yields four equations that can be solved for the four unknowns. The resulting reflection coefficient from the slab can be ex-

* mansuripur@physics.harvard.edu

pressed in terms of the single-surface Fresnel coefficients as

$$r \equiv \frac{E_1^L}{E_1^R} = \frac{r_{12} + r_{23}\exp(2ik_{2z}^R d)}{1-\nu} \tag{4}$$

where

$$\nu = r_{21}r_{23}\exp(2ik_{2z}^R d) \tag{5}$$

is referred to as the roundtrip coefficient; the amplitude of a planewave circulating in the slab is multiplied by this factor after each roundtrip in the absence of any sources outside the slab. (Although we explicitly discuss s-polarized light, our conclusions as well as Eqs. 4 and 5 hold for both polarization states.) We emphasize that the reflection coefficient given by Eq. 4 is a valid solution to Maxwell's equations for any value of ν. The roundtrip coefficient ν has an important physical meaning, and intuitively one would expect three different regimes of behavior when the magnitude of ν is less than, equal to, or greater than one. The case where $|\nu| < 1$ governs passive slabs (in most but not all cases) and sufficiently weakly amplifying slabs. When $\nu = 1$ the slab behaves as a laser and emits light even in the absence of an incident wave, which manifests itself mathematically as an infinitely large reflection amplitude. The case where $|\nu| > 1$, however, has received scant attention in the literature [4, 15].

Perhaps the reason for the neglect of the $|\nu| > 1$ steady-state solution is the seemingly intuitive assumption that there cannot be a steady-state solution when $|\nu| > 1$–due to gain saturation in a laser, for instance. (See supplementary information.) This assumption is only reinforced by examining a second well-known solution method for the reflection coefficient that decomposes the reflected wave amplitude E_1^L into a sum over partial waves, yielding the reflection coefficient

$$r = r_{12} + t_{12}t_{21}r_{23}\exp(2ik_{2z}^R d)\sum_{m=0}^{\infty}\nu^m. \tag{6}$$

Heuristically, the first term r_{12} (hereinafter referred to as the "specular" partial wave) of Eq. 6 results from the single-surface reflection of the incident wave at the 1-2 interface, and the geometric series accounts for the contributions to the reflected wave following multiple roundtrips within the slab. When $|\nu| < 1$, the geometric series in Eq. 6 converges to $(1-\nu)^{-1}$, giving the same result as found by matching the boundary conditions in Eq. 4. When $|\nu| > 1$, however, the geometric series diverges and the reflection coefficient is infinite. Intuitively, this divergence seems reasonable, since we expect any light that couples into a slab with $|\nu| > 1$ to be amplified after each roundtrip, and therefore grow without bound. Nevertheless, Eq. 4 yields a finite reflection coefficient even when $|\nu| > 1$, so how can we reconcile these two very different solutions?

In fact, the partial wave method *can* be used to find the $|\nu| > 1$ convergent solution. First, note that the reflection coefficient given by Eq. 4 is invariant under the transformation $k_{2z}^R \leftrightarrows k_{2z}^L$. (This is not surprising as it can be interpreted simply as a relabeling of the waves $E_2^R \leftrightarrows E_2^L$ in Eq. 3 that does not affect the final result.) Applying this same transformation to the partial wave sum of Eq. 6 [15], we can express the reflection coefficient as

$$r = r'_{12} + t'_{12}t'_{21}r'_{23}\exp(2ik_{2z}^L d)\sum_{m=0}^{\infty}\nu'^m, \tag{7}$$

where the prime indicates the substitution $k_{2z}^R \to k_{2z}^L$. Because the new roundtrip coefficient, $\nu' = r'_{21}r'_{23}\exp(2ik_{2z}^L d)$, is equal to ν^{-1}, in cases where $|\nu| > 1$ the primed partial wave sum of Eq. 7 will converge to the reflection coefficient of Eq. 4. (Incidentally, this duality between ν and ν^{-1} provides a simple mathematical explanation for the loss/gain duality observed by others [4–8]).

The physical implications of the substitution $k_{2z}^R \to k_{2z}^L$ in the partial wave sum can best be seen by examining the behavior of a "finite-diameter" beam of light incident obliquely on the slab. By numerically superposing a finite number of planewave solutions to Eq. 3 with appropriate amplitudes and incidence angles in the range $27.47° < \theta < 32.53°$ (see supplementary information), we create a Gaussian (to within the sampling accuracy) beam incident on the slab at 30° with a full-width at half-maximum beam-diameter of 13.3 μm. All media are nonmagnetic, and we choose $\epsilon_1 = \epsilon_3 = 2.25$ and the slab to be an amplifying medium with $\epsilon_2 = 1 - 0.01i$. The free-space wavelength of the beam is $\lambda_o = 1$ μm. We can examine the transition at $|\nu| = 1$ simply by varying d, since both $|r_{21}|$ and $|r_{23}|$ are less than one (and independent of d), whereas $|\exp(2ik_{2z}^R d)|$ (and hence ν) increases monotonically with d (because k_{2z}^R has a negative imaginary part). A plot of the field $E_y(x, z)$ at one instant of time is shown in Fig. 2(a) for $d = 19$ μm, which was chosen so that $|\nu|$ is slightly less than one for all constituent planewaves of the beam $(0.46 < |\nu| < 0.99)$. The arrows overlying the plot point in the direction of the time-averaged Poynting vector within their vicinity, indicating the direction of energy flow in the system, and the incident beam is uniquely identified by the white arrow. The beam behaves as we expect it to: the incident beam strikes the slab near $(x = 0, z = 0)$, giving rise to a specularly reflected beam as well as a refracted beam that 'zig-zags' up the slab, which in turn generates a reflected beam in medium one each time it strikes the 2-1 interface. (The field amplitude is plotted on a linear scale, and so the incident beam as well as the specularly reflected beam appear faint relative to the subsequently amplified portions of the beam.) Each of these reflected beams can intuitively be associated with a term of the partial wave expansion of Eq. 6–either the specular term or the mth term of the geometric series.

In Fig. 2(b) all parameters are kept the same except the slab thickness is increased to $d = 28$ μm, resulting in $|\nu| > 1$ for all constituent planewaves of the Gaussian

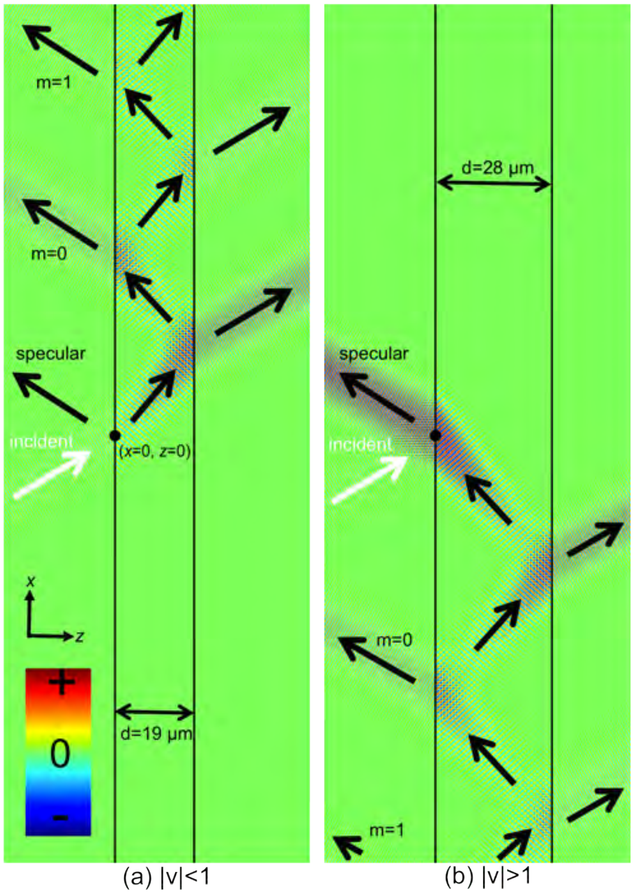

FIG. 2. Plots of the field $E_y(x, z)$ at one instant of time for a Gaussian beam (indicated with the white arrow) incident on an amplifying slab for which (a) $|\nu| < 1$ ($d = 19$ μm) and (b) $|\nu| > 1$ ($d = 28$ μm). The black dot indicates the origin of the coordinate system.

beam ($1.01 < |\nu| < 2.58$). Based solely on the plot of the field amplitude and not on the direction of energy flow indicated by the arrows, it may appear that the incident beam strikes the interface and negatively refracts in the slab, then zig-zags downwards in the $-\hat{\boldsymbol{x}}$ direction, giving rise to many reflected beams in medium one (and transmitted beams in medium three) which emanate from points on the slab with $x < 0$. Such an explanation was offered for simulations similar to ours [12, 13] to attempt to justify negative refraction in an active, nonmagnetic medium. However, by analyzing the Poynting vector we see that the energy in the beam zig-zags *up* the slab, so this phenomenon is distinct from negative refraction, despite the similarity in the positions of the reflected and transmitted beams. (In the supplementary information, a video of the time-dependent behavior of a "finite-duration" pulse of light more clearly illustrates that energy flows in the $+x$-direction.) We will refer to the field in the slab at $x < 0$ as the "pre-excitation," so-called because it occurs before the central lobe of the incident beam arrives at the slab. (We will discuss possible mechanisms for the origin of the pre-excitation shortly, but for now simply accept it as a necessary consequence of the Fresnel solution.) Each reflected beam in Fig. 2(b) can be associated either with the specular term r'_{12} or with the mth term of the primed partial wave expansion in Eq. 7. In particular, note that $r'_{12} = r_{12}^{-1}$; since $|r_{12}| < 1$ (in most cases of practical interest), this means that $|r'_{12}| > 1$, and so the primed partial wave expansion mathematically predicts the amplification of the specularly reflected beam only when $|\nu| > 1$. From Fig. 2(b) we see that this amplification occurs because the specular beam receives energy from the transmission of the pre-excited field through the 2-1 interface.

It is interesting to compare a lossy and an amplifying slab in the limit as $d \to \infty$. In a lossy slab (for which $\text{Im}(k_{2z}^R) > 0$), the roundtrip coefficient $\nu \to 0$ as $d \to \infty$, and so the reflection coefficient r approaches the single-surface solution r_{12}, as expected, because the geometric series in Eq. 6 makes no contribution. In a gainy slab (for which $\text{Im}(k_{2z}^R) < 0$), $\nu \to \infty$ as $d \to \infty$, but $\nu' \to 0$ and so we see from Eq. 7 that $r \to r'_{12}$ (which means that the field in the slab is dominated by the wavevector k_{2z}^L, i.e., $E_2^R/E_2^L \to 0$). The reason the limiting treatment of $d \to \infty$ for a gainy slab does not yield the proper single-surface reflection coefficient is that no matter how large one chooses to make d, the nonzero reflection r_{23} at the back-facet of the slab allows for the amplification of the pre-excited field; for $|\nu| \gg 1$, this results in the left-propagating wavevector k_{2z}^L dominating the behavior of the slab while the amplitude of the wave associated with k_{2z}^R diminishes substantially, and the reflection coefficient correspondingly approaches r'_{12}. Nevertheless, the right-propagating wave is essential in spite of its seemingly inconsequential amplitude, as it is responsible for generating the left-propagating wave by way of the back-facet reflection. In the case of two truly semi-infinite media (i.e., media one and two), the absence of a back-facet prevents any roundtrip amplification of the pre-excitation, so the only wavevector that exists in the transmission medium is k_{2z}^R–as agreed upon in all cases except possibly amplified TIR–and the single-surface reflection coefficient is correctly given by r_{12}, not r'_{12}.

The scenario under scrutiny in the "amplified TIR" problem concerns light in a high-index passive medium incident above the critical angle for TIR on a semi-infinite lower-index amplifying medium. Some have argued that the incident wave excites the wavevector k_{2z}^L in the transmission medium rather than the usual k_{2z}^R, resulting in the reflection coefficient r'_{21} and an accompanying amplified specular reflection [14–17, 19]. Before considering an amplifying medium of semi-infinite thickness, however, let us consider a finite-thickness amplifying slab and understand the role that ν plays in this problem. For the same parameters as those used in Fig. 2(b), $|\nu|$ increases monotonically with increasing incidence angle θ for s-polarized light; in particular, $|\nu|$ exceeds one as long as $\theta > 27.43°$. As θ approaches and surpasses the critical angle for TIR, $\theta_c = 41.8°$, $|\nu|$ quickly becomes extremely large due to the negatively increasing $\mathrm{Im}(k_{2z}^R)$. (For $\theta = 41°$, $|\nu| = 9.34 \times 10^3$, and for $\theta = 42°$, $|\nu| = 1.40 \times 10^{15}$.) More generally, TIR from a gainy slab is well within the regime $|\nu| >> 1$ (for any reasonable thickness d). This is then comparable to the case of large d, for which we demonstrated in the previous paragraph that the existence of the left-propagating k_{2z}^L relies on the nonzero back-facet reflection r_{23} [18], resulting in a reflection coefficient approaching r'_{12}. When the amplifying medium is semi-infinite, however, $\nu = 0$ because r_{23} is effectively zero, therefore the pre-excitation does not occur and consequently the wave with wavevector k_{2z}^L does not exist. This suggests that k_{2z}^R is the correct choice for the transmitted wavevector even in the case of TIR from an amplifying medium.

The Fresnel solution for a slab with $|\nu| > 1$ is a steady-state harmonic solution, in the sense that if the field distribution presented in Fig. 2(b) exists at time t_o, then as time is evolved forward the field at each point in space will vary harmonically with frequency ω. We have not yet addressed the question of how the pre-excited field is established in the first place, beginning from a state with no electromagnetic fields anywhere in space. One possibility that does not violate causality is to consider that the Gaussian beam does not have a truly finite spatial width, but rather a rapidly decaying "side-tail" in the direction normal to the propagation direction. The side-tail is capable of injecting a small amount of energy into the slab at positions $x \ll 0$. When $|\nu| > 1$, light in the slab gains more during one roundtrip than it loses to transmission at both facets, and so this initially small amount of energy is amplified, resulting in the pre-excited field. For a beam or pulse at normal incidence, the leading edge (also known as the Sommerfeld front-runner) can perhaps play a similar role. The key point is that when $|\nu| > 1$, our intuition about the arrival time and arrival position of the beam (or pulse) misleads us because amplification by the slab acts on typically negligible field amplitudes to dramatically alter the character of the field. Finite-difference time-domain methods, applied to a slab whose dispersion obeys the Kramers-Kronig relations, are perhaps best suited to investigate the causal evolution of the pre-excitation. We note also that the Fresnel formalism, by implicitly beginning with the time-harmonic subset of Maxwell's equations, can only elucidate the non-divergent solutions to the full time-dependent equations. There can certainly exist divergent solutions, as demonstrated by finite-element simulations of a wave with a well-defined start-time incident normally on a slab with $|\nu| > 1$ [4]. Therefore, it is worth questioning the stability of the convergent solution: can it be seen in practice, or will it be obscured by other solutions? Some may also question whether a cavity with net roundtrip gain can exist in the steady-state in the first place; we note that the experimental work already done on the amplification of evanescent waves [14, 20, 21] (a regime for which $|\nu| > 1$) has not detected any instabilities, due to spontaneous emission or otherwise. A laser with destructive external self-feedback is also a stable system for which $|\nu| > 1$, which we discuss in the supplementary information. Finally, we emphasize that $|\nu|$ can exceed one even in passive media provided $|r_{21}r_{23}|$ exceeds one, which can happen for incident evanescent waves near surface plasmon resonances and therefore provides insight to the behavior of the negative-index lens [1–3]. This will be treated fully in future work, but we mention it here to prevent the reader from simply dismissing the $|\nu| > 1$ scenario as an unphysical mathematical peculiarity of gain media. In the end, only an experiment can determine whether the solution we have presented is a physical one; at the very least, we hope that our application of the Fresnel formalism to the unintuitive case of a cavity whose roundtrip gain exceeds the loss has offered fresh perspective on a few persistent controversies, and perhaps generated interesting ideas for future experiments.

The pulse simulation was run on the Odyssey cluster supported by the Harvard FAS Research Computing Group. TSM is supported by an NSF Graduate Research Fellowship. We thank Alexey Belyanin and Steven Byrnes for helpful discussions.

---

[1] J.B. Pendry. "Negative refraction makes a perfect lens." Phys. Rev. Lett. **85**, 3966-69 (2000).

[2] D.R. Smith, D. Schuring, M. Rosenbluth, S. Schultz, S.A. Ramakrishna, and J.B. Pendry. "Limitations on subdiffraction imaging with a negative refractive index slab." Appl. Phys. Lett. **82**, 1506-1508 (2003).

[3] N. Garcia and M. Nieto-Vesperinas. "Left-Handed Materials Do Not Make a Perfect Lens." Phys. Rev Lett. **88**, 20, 207403 (2002).

[4] X. Jiang, Q. Li and C.M. Soukoulis. "Symmetry between absorption and amplification in disordered media." Phys. Rev. B **19**, 9007 (1999).

[5] S.A. Ramakrishna. "On the dual symmetry between absorbing and amplifying random media." Pramana **62**, 1273-1279, (2004).

[6] J.C.J. Paasschens, T. Sh. Misirpashaev and C.W.J. Beenakker. "Localization of light: dual symmetry between absorption and amplification." Phys. Rev. B **54**, 11887 (1996).

[7] P. Pradhan and N. Kumar. "Localization of light in coherently amplifying random media." Phys. Rev. B **50**, 9644 (1994).

[8] Z.Q. Zhang. "Light amplification and localization in randomly layered media with gain." Phys. Rev. B **52**, 7960 (1995).

[9] Y. Chen, P. Fischer, and F.W. Wise. "Negative Refraction at Optical Frequencies in Nonmagnetic Two-Component Molecular Media." Phys. Rev. Lett. **95**, (2005).

[10] S.A. Ramakrishna. "Comment on 'Negative Refraction at Optical Frequencies in Nonmagnetic Two-Component Molecular Media.'" Phys. Rev. Lett. **98**, (2007).

[11] S.A. Ramakrishna and O.J.F. Martin. "Resolving the wave vector in negative index media." Opt. Lett. **30**, 19, 2626-2628 (2005).
[12] J.B. Geddes III, T.G. Mackay, and A. Lakhtakia. "On the refractive index for a nonmagnetic two-component medium: Resolution of a controversy." Opt. Comm. **280**, 120-125 (2007).
[13] B. Nistad and J. Skaar. "Causality and elecromagnetic properties of active media." Phys. Rev. E **78**, 036603 (2008).
[14] C.J. Koester. "Laser action by enhanced total internal reflection." IEEE J. Quantum Electron. **QE-2**, 580-4 (1966)
[15] P.R. Callary and C.K. Carniglia. "Internal reflection from an amplifying layer." J. Opt. Sci. Am. **66**, 8 (1976).
[16] J. Fan, A. Dogariu, and L.J. Wang. "Amplified total internal reflection." Opt. Express **11**, 299-308 (2003).
[17] K.J. Willis, J.B. Schneider, and S.C. Hagness. "Amplified total internal reflection: theory, analysis, and demonstration of existence via FDTD." Opt. Express **16**, 1903-14 (2008).
[18] A. Siegman. "Fresnel Reflection and Evanescent Gain." OPN, 38–45 (Jan. 2010).
[19] J.O. Grepstad and J. Skaar. "Total internal reflection and evanescent gain." Opt. Express **19**, 21404 (2011).
[20] E.P. Ippen and C.V. Shank. "Evanescent-field-pumped dye laser." Appl. Phys. Lett. **21**, 301 (1972).
[21] H.H. Choi, H.J. Kim, J. Noh and W. Jhe. "Amplified total internal reflection from a plane dielectric boundary." Jour. Korean Phys. Soc. **42**, 80 (2003).

Supplementary information for version 3 (submitted to PRL 8/23/2013).
Changes relative to version 2 are highlighted.

Fresnel reflection from a cavity with net roundtrip gain: Supplementary Information

Tobias S. Mansuripur[1,*] and Masud Mansuripur[2]

[1]*Department of Physics, Harvard University, Cambridge, MA 02139*

[2]*College of Optical Sciences, The University of Arizona, Tucson, AZ 85721*

[*] mansuripur@physics.harvard.edu

I. PRESCRIPTION FOR R AND L SUPERSCRIPTS

The energy flux of an s-polarized planewave whose E-field is given by

$$E_y(x, z, t) = E_0 \exp(ik_x x + ik_z z - i\omega t) \tag{1}$$

in a medium (ϵ, μ) is given by the time-average of the Poynting vector $\boldsymbol{S} \equiv \vec{E} \times \vec{H}$,

$$\langle \vec{S} \rangle = \frac{|E_0|^2}{2\omega\mu_0} e^{-2\mathrm{Im}(k_z)z} \left(\mathrm{Re}\left[\frac{k_x}{\mu}\right] \hat{x} + \mathrm{Re}\left[\frac{k_z}{\mu}\right] \hat{z} \right). \tag{2}$$

Therefore, energy flows in the $+z$-direction ('to the right,' in our convention) when $\mathrm{Re}(k_z/\mu) > 0$, and we denote the wavevector k_z which satisfies this condition with the superscript R. Any value k_z for which $\mathrm{Re}(k_z/\mu) < 0$ is accordingly labeled with a superscript L. It follows from these definitions that k_z^R must make an acute angle with μ in the complex plane. (For p-polarized light energy flows in the $+z$-direction when $\mathrm{Re}(k_z/\epsilon) > 0$, and so k_z^R makes an acute angle with ϵ in the complex plane.)

In cases where $\langle S_z \rangle = 0$, we must establish a prescription for resolving the ambiguity in the choice of superscript, which is best illustrated by an example. Consider the case where medium one is a lossless dielectric ($\epsilon_1 > 1$, $\mu_1 = 1$), medium two is vacuum, and the incident propagating wave satisfies $k_x > k_0$, where $k_0 \equiv \omega/c$, so that the two choices for k_{2z} are $\pm i\sqrt{k_x^2 - k_0^2}$. Both choices for k_{2z} yield pure evanescent waves and carry no energy along the z-direction. By adding a small amount of loss to medium two, so that $\epsilon_2 = 1 + i\epsilon_2''$ where $\epsilon_2'' > 0$, the two choices for k_{2z} deviate slightly from the imaginary axis as shown in Fig. S1(a). Now both waves carry non-zero energy along the z-direction; the first quadrant solution is k_{2z}^R (which can be seen quickly because it makes an acute angle with $\mu_2 = 1$) and the third quadrant solution is k_{2z}^L. Our prescription to establish k_{2z}^R for a true vacuum (i.e., $\epsilon_2'' = 0$) is to take the limit $\epsilon_2'' \to 0$, which yields k_{2z}^R as the solution along the positive imaginary axis.

Beware that if one adds a small amount of gain rather than loss to medium two, so that $\epsilon_2 = 1 + i\epsilon_2''$ where $\epsilon_2'' < 0$, then the two solutions for k_{2z} exist in the second and fourth quadrants as shown in Fig. S1(b), and in this case k_{2z}^R points predominantly along the *negative* imaginary axis. Thus, we see that the two limiting cases as gain or loss approaches zero do not yield the same result:

$$\lim_{\epsilon_2'' \to 0^+} k_{2z}^R = - \lim_{\epsilon_2'' \to 0^-} k_{2z}^R. \tag{3}$$

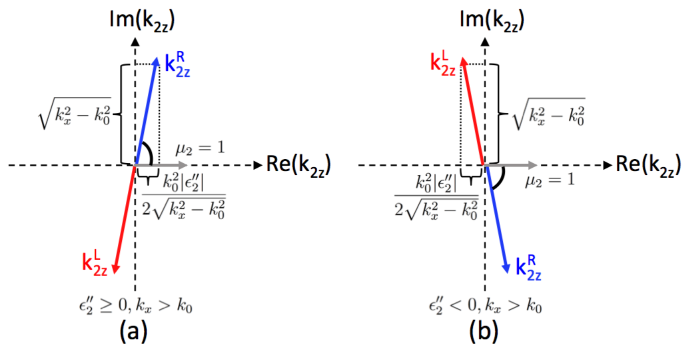

FIG. S1. Choosing the R or L label for an evanescent wave. (a) The two choices for k_{2z} are shown for the case of a slightly "lossy vacuum" ($\epsilon_2 = 1 + i\epsilon_2''$ where $\epsilon_2'' > 0$, $\mu_2 = 1$), for the case $k_x > k_0$. The first quadrant solution carries energy to the right and is labeled k_{2z}^R, and our prescription is to take the limit $\epsilon_2'' \to 0$ to determine that k_{2z}^R in the lossless case is along the positive imaginary axis. (b) For a slightly "gainy vacuum" ($\epsilon_2'' < 0$) the two solutions for k_{2z} are in the second and fourth quadrants, and k_{2z}^R approaches the negative imaginary axis as $\epsilon_2'' \to 0$. The magnitudes of the real and imaginary parts of k_{2z} in (a) and (b) are approximated using the first order Taylor expansion for small ϵ_2'': $k_0^2|\epsilon_2''| \ll k_x^2 - k_0^2$.

To have an unambiguous labeling convention for the case $\epsilon_2'' = 0$, we emphasize that one must take the limit as *loss* approaches zero, which can be different from the limit as gain approaches zero in the case of evanescent waves.

Finally, it is worth noting that this discontinuity in the two limiting cases, apart from being a footnote in establishing a labeling convention, is actually at the heart of the debate over single-surface amplified TIR. When medium two has gain, if one chooses k_{2z}^R as the transmitted wavevector (in accordance with our postulate), then it seems unphysical that as the gain approaches (but does not reach) zero the transmitted wave should still be strongly amplified. To remedy this situation it has been suggested that the correct choice for the transmitted wavevector should be k_{2z}^L when medium two has gain and $k_x > k_0$, so that the transmitted wave decays in the $+z$-direction. We believe, instead, that the discontinuity in the two limits is not as unphysical as it might appear at first: the transmitted wave propagates a large distance in the x-direction while barely moving forward in the z-direction (since $k_x \gg \mathrm{Re}(k_{2z}^R)$), so the large gain in the z-direction is actually a result of the long propagation distance along the x-direction. Far more unphysical, in our opinion, is the decision to switch the transmitted wavevector from k_{2z}^R when $k_x < k_0$ to k_{2z}^L when $k_x > k_0$.

All of these arguments aside, however, the purpose of our paper has been to demonstrate the Fresnel mechanism by which the specularly reflected beam from a finite-thickness slab is amplified, both below and above the critical angle.

II. GAIN SATURATION

Physicists familiar with the principles of lasers should be rightfully wary of a steady-state solution with roundtrip coefficient $|\nu|$ greater than one. When an active medium is pumped strongly enough to generate a sufficiently large population inversion to yield $|\nu|$ greater than one, light initially generated by spontaneous emission in the cavity will be amplified after each roundtrip. However, the field amplitude does not grow without bound–as the field gains strength the upper state lifetime is reduced by stimulated emission, which causes the population inversion to decrease to a level such that $\nu = 1$, resulting in steady-state lasing. This gain reduction with increasing field amplitude is known as gain saturation. In a laser, therefore, the situation $|\nu| > 1$ is only a transient state. It clearly cannot be a steady-state solution, because the field would grow without bound.

The situation changes when we allow an incident wave to strike the active medium, as we do in this paper. Note that ν is defined as the roundtrip coefficient *in the absence of an incident wave*; that is, the reflectivity r_{21} is calculated by assuming that there is no wave in medium one arriving at the cavity. To account for the incident wave, we can define an effective facet reflectivity at the two-one interface $r_{21}^{\text{eff}} \equiv E_2^R/E_2^L$. Furthermore, we can define an effective roundtrip coefficient in the slab which replaces r_{21} with r_{21}^{eff}, that is, $\nu^{\text{eff}} = r_{21}^{\text{eff}} r_{23} \exp(2ik_{2z}^R d)$. We emphasize that *every possible steady-state solution to the problem under consideration, whether the slab is passive or active, and whether there is an incident wave or not, satisfies the condition* $\nu^{\text{eff}} = 1$. This is a fundamental property of steady-state solutions: the field in the slab must regenerate itself after every roundtrip, once all sources and sinks have been accounted for. Therefore, in situations where $|\nu| > 1$, the incident wave must, upon transmission into medium two, interfere destructively with the circulating field in the slab so that $|r_{21}^{\text{eff}}| < |r_{21}|$, which ultimately forces ν^{eff} toward 1. In summary, when there is no incident wave the situation $|\nu| > 1$ is temporary because the field will grow until gain saturation (a nonlinear effect) forces the $\nu = 1$ solution. With an incident wave, a linear steady-state solution is possible even when $|\nu| > 1$ because of the

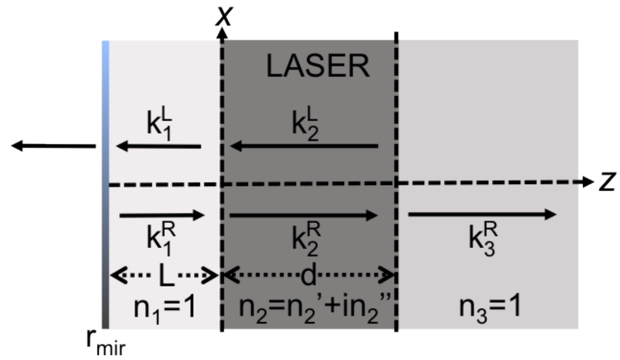

FIG. S2. Schematic of a laser with self-feedback. A mirror of reflectivity r_{mir} allows light emitted by medium two to be re-injected. The feedback alters the laser threshold, and can be modeled by an effective facet reflectivity r_{21}^{eff}, which differs from the bare facet reflectivity r_{21} in the absence of feedback.

reduction in the effective facet reflectivity r_{21}^{eff}, which prevents the unbounded growth of the fields so that one does not have to rely on gain saturation to avoid a nonphysical divergence.

An example that illustrates well the distinction between ν and ν^{eff} is a laser subject to self-feedback, shown in Fig. S2. In this case, the laser (medium 2) emits light in a direction normal to both of its facets, and a mirror of reflectivity r_{mir} placed a distance L from the left facet reflects some of the laser output and creates the wavevector $\mathbf{k}_1^R$, which is then reinjected into the laser. (To be clear, this is different from the situation considered in our paper, for which $\mathbf{k}_1^R$ is generated by an external source.) For a suitable choice of L and r_{mir}, the re-injected wave will interfere destructively with the circulating field in the gain medium, effectively reducing the reflection coefficient at the 2-1 interface without altering its phase. Specifically, for $n_1 = 1$ and $n_2 = n_2' + in_2''$, the facet reflectivity in the absence of feedback is $r_{21} = (n_2 - 1)/(n_2 + 1)$. The feedback from the external mirror will reduce the effective reflection coefficient to $r_{21}^{\text{eff}} = \alpha r_{21}$, where α is real and satisfies $0 \leq \alpha \leq 1$, provided r_m and L are chosen such that

$$r_m \exp(i4\pi L/\lambda_o) = \frac{(1-\alpha)r_{21}}{\alpha r_{21}^2 - 1}. \tag{4}$$

So long as $|r_{21}| < 1$ (which is generally the case), it is possible to choose r_m and L so that $|r_m| < 1$, i.e., the external mirror is a simple, passive component. As an application of Eq. 4, consider the case $n_2 = 1.5 - 0.01i$ and $\lambda_o = 1$ μm. The bare facet reflectivity is $r_{21} = 0.2\exp(-0.016i)$. If we want the effective facet reflectivity to be $r_{21}^{\text{eff}} = 0.1\exp(-0.016i)$

(half the magnitude but the same phase as the bare facet reflectivity), we choose $\alpha = 0.5$ and find that we should place a mirror of reflectivity $r_m = -0.102$ a distance $L = 0.499$ μm from the left facet.

Because the external mirror reduces the effective facet reflectivity so that $|r_{21}^{\text{eff}}| < |r_{21}|$, the effective roundtrip coefficient $|\nu^{\text{eff}}|$ will be less than $|\nu|$. When medium two is pumped beyond the lasing threshold, ν^{eff} will be clamped to 1, which means that $|\nu|$ will be greater than 1. There also exists a subthreshold pumping regime for which $|\nu| > 1$. The example of the laser with self-feedback demonstrates that an active cavity subject to an incident field can have a roundtrip coefficient ν with magnitude greater than 1, and that this situation is neither transient nor unstable.

In the above example, there is always a well-defined phase relationship between the field circulating in the laser and the re-injected field. This will not necessarily be true when the incident wave is generated by an external source: spontaneous emission, being a stochastic process, can give rise to a field within medium two that has no well-defined phase relationship with the incident wave. Whether this leads to instabilities remains to be decisively answered, but we note that the existing experimental evidence in the $|\nu| > 1$ regime has not, to our knowledge, detected any instabilities [14, 20, 21]. It is also worth commenting on an important difference between spontaneously-emitted cavity-photons traveling parallel to the surface-normal of the cavity facets versus those emitted at an oblique angle to the optical axis. For photons emitted parallel to the surface-normal, the slab is clearly a cavity that provides feedback, as the light retraces its path on every roundtrip. However, as mentioned in the main text, the roundtrip coefficient is a function of incidence angle, and for the material parameters used for Fig. 2(b), $|\nu|$ exceeds one only for incidence angles $\theta > 27.43°$. In particular, this means that $|\nu| < 1$ for $\theta = 0$; therefore, the gain is insufficient for a spontaneously emitted cavity photon in the $\theta = 0$ direction to cause lasing. In contrast, a cavity photon emitted spontaneously at a large θ such that $|\nu| > 1$ would seem to experience net amplification after each roundtrip. (We say 'seem to' because this is the intuitive interpretation to us; however, we mention again that the experiments that have probed the regime of $|\nu| > 1$ have not detected such problems with spontaneous emission [14, 20, 21].) Should these obliquely-traveling spontaneously-emitted photons prove problematic in a future experiment, however, they will anyway exit the slab at the top and bottom facets, since any real slab must have a finite length in the x-direction. One could also coat the top and bottom facets with broadband antireflection coatings to facilitate this removal; this way, these spontaneously emitted photons would leave the slab before they are amplified to the point where they saturate the gain. We mention this only as a consideration for a potential experiment that looks for the pre-excitation mechanism. Our intent in this Letter has been to explore the predictions of the Fresnel solution for a slab with $|\nu| > 1$; in the end, only experiments can decide whether this solution is physical.

III. PULSE OF LIGHT INCIDENT ON GAINY SLAB WITH $|\nu| > 1$

The video file pulse_video.avi included online is a time-lapse video of $|E_y|^2$ of a pulse of light, rather than a beam, for the same material parameters as in Fig. 2(b) of the main text: $\epsilon_1 = \epsilon_3 = 2.25$, $\epsilon_2 = 1 - 0.01i$, $\mu_1 = \mu_2 = \mu_3 = 1$, $d = 28$ μm. The white vertical lines in the video identify the 1-2 and 2-3 interfaces. The incident pulse is s-polarized and Gaussian in both space (FWHM = 13.3 μm) and time (FWHM = 50 fs, or 15 optical cycles). The central wavelength of the pulse is $\lambda_o = 1$ μm, and the mean incidence angle (i.e., averaged over all constituent planewaves) is 30°. The size of each video frame is 210 μm by 150 μm (height by width). The time elapsed between frames is 10 fs, and the entire video spans 1.22 ps (123 frames total). The field $|E_y|^2$ is plotted on a logarithmic scale covering 3 decades, i.e. red corresponds to the maximum intensity and blue corresponds to intensities less than or equal to 1/1000th of the maximum. The background in this image is blue, which corresponds to the minimum of $|E_y|^2$, whereas the background in Figs. 2(a) and 2(b) is green because it is the field E_y that was plotted in that case, so that blue corresponded to the maximum negative field.

In the video, one first sees the incident pulse near the bottom left of the screen, traveling up and to the right. The pre-excitation is soon seen in the slab at the bottom of the frame, and the reflected pulse that corresponds to the $m = 1$ term in the primed partial wave expansion leaves the slab and propagates up and to the left in medium one. The pre-excitation in the slab then undergoes one roundtrip as it zig-zags upward, giving rise to a transmitted pulse in medium three followed by the $m = 0$ reflected pulse in medium one. The pre-excitation then makes one more roundtrip, giving rise to another transmitted pulse in medium three, and then approaches the two-one interface at the same time the incident pulse arrives from the opposite side. The two pulses interfere in such a way as to yield an amplified specularly reflected pulse by entirely depleting the energy content of the slab. The fact that the pre-excitation in the slab travels in the $+x$-direction clearly distinguishes this behavior from negative refraction.

IV. DESCRIPTION OF SIMULATION

The E-field plots of the Gaussian beams and the video of the pulse were created using MATLAB. The field at each pixel is determined by superposing a large (but of course finite) number of planewave solutions. Therefore, the plots represent analytical solutions to Maxwell's equations.

As described in the main text, the response of the slab to an incident s-polarized planewave with amplitude E_1^R and wavevector $\boldsymbol{k}_1^R = k_x\hat{\boldsymbol{x}} + k_{1z}^R\hat{\boldsymbol{z}}$ is given by

$$E_y(x,z) = \begin{cases} E_1^R \exp(ik_x x + ik_{1z}^R z) + E_1^L \exp(ik_x x + ik_{1z}^L z) & : z \leq 0 \\ E_2^R \exp(ik_x x + ik_{2z}^R z) + E_2^L \exp(ik_x x + ik_{2z}^L z) & : 0 \leq z \leq d \\ E_3^R \exp[ik_x x + ik_{3z}^R (z-d)] & : z \geq d \end{cases} \tag{5}$$

and the time-dependence factor $\exp(-i\omega t)$ is not explicitly written. The wavevector components k_{2z}^R and k_{3z}^R are determined by the dispersion relation

$$k_{\ell z}^R = \sqrt{(\omega/c)^2\mu_\ell\epsilon_\ell - k_x^2}, \tag{6}$$

where μ_ℓ and ϵ_ℓ are the relative magnetic permeability and electric permittivity constants of material ℓ, and the sign of the square root is chosen according to the prescription described in Supplementary Sec. 1. The four unknown wave amplitudes are found by satisfying Maxwell's boundary conditions to be

$$E_2^R = \frac{2k_{1z}^R(k_{3z}^R + k_{2z}^R)E_1^R}{(k_{2z}^R + k_{1z}^R)(k_{3z}^R + k_{2z}^R) + \exp(2ik_{2z}^R d)(k_{3z}^R - k_{2z}^R)(k_{2z}^R - k_{1z}^R)} \tag{7}$$

$$E_2^L = \frac{-2k_{1z}^R(k_{3z}^R - k_{2z}^R)E_1^R}{(k_{2z}^R - k_{1z}^R)(k_{3z}^R - k_{2z}^R) + \exp(-2ik_{2z}^R d)(k_{3z}^R + k_{2z}^R)(k_{2z}^R + k_{1z}^R)} \tag{8}$$

$$E_1^L = E_2^R + E_2^L - E_1^R \tag{9}$$

$$E_3^R = E_2^R \exp(ik_{2z}^R d) + E_2^L \exp(-ik_{2z}^R d). \tag{10}$$

To construct the Gaussian beam from the planewave solutions, we begin by expressing E_y in the $z = 0$ plane for a beam traveling parallel to the z-axis

$$E_y(x, z = 0) = E_0 \exp\left(-\frac{x^2}{2\sigma_x^2}\right), \tag{11}$$

where E_0 is the peak amplitude and σ_x is directly proportional to the spatial FWHM

$$w_x = 2\sqrt{2\ln 2}\sigma_x. \tag{12}$$

By Fourier transforming and subsequently inverting the transform, the field can equivalently be written as an integral in k-space,

$$E_y(x, z = 0) = \int_{-\infty}^{\infty} dk_x E_1^R(k_x)\exp(ik_x x), \tag{13}$$

where

$$E_1^R(k_x) = \frac{E_0\sigma_x}{\sqrt{2\pi}}\exp\left(\frac{-k_x^2}{2(1/\sigma_x)^2}\right), \tag{14}$$

and the FWHM in k-space is

$$w_k = 2\sqrt{2\ln 2}/\sigma_x. \tag{15}$$

To propagate the beam beyond the $z = 0$ plane, we associate with each value of k_x a component k_{1z}^R such that the total wavevector obeys the dispersion relation in medium one,

$$k_{1z}^R(k_x) = \sqrt{(\omega/c)^2\mu_1\epsilon_1 - k_x^2}. \tag{16}$$

Now the Gaussian beam can be expressed as a function of x and z by

$$E_y(x, z) = \int_{-\infty}^{\infty} dk_x E_1^R(k_x)\exp[i(k_x x + k_{1z}^R z)]. \tag{17}$$

At this point, we must approximate the integral in Eq. 17 by discretization so that the calculation can be carried out by a computer. We restrict k_x to a finite sampling width w_s given by $-w_s/2 \leq k_x \leq w_s/2$, and sample the beam equidistantly within this region with a total number of samples N_s. The integral in Eq. 17 is approximated by the sum

$$E_y(x, z) = \sum_{k_x=-w_s/2}^{w_s/2} \Delta k_x E_1^R(k_x)\exp[i(k_x x + k_{1z}^R z)], \tag{18}$$

where

$$\Delta k_x = \frac{w_s}{N_s - 1}. \tag{19}$$

At this point, it is helpful to think of E_1^R, k_x, and k_{1z}^R as vectors containing N_s numerical elements each. To rotate the beam so that it travels at an angle θ to the z-axis, we perform the transformation

$$k_x \rightarrow \cos(\theta)k_x + \sin(\theta)k_z \tag{20}$$

$$k_{1z}^R \rightarrow -\sin(\theta)k_x + \cos(\theta)k_{1z}^R \tag{21}$$

on each element of k_x and k_{1z}^R. (The Fourier amplitude of each plane-wave $E_1^R(k_x)$ is unaffected by the rotation in the case of s-polarized light.) Finally, to displace the waist of the beam to some location (x_0, z_0) in the incidence medium, one must multiply each Fourier amplitude by

$$E_1^R(k_x) \to E_1^R(k_x) \exp[-i(k_x x_0 + k_{1z}^R z_0)]. \tag{22}$$

With these redefined values for E_1^R, k_x, and k_{1z}^R, the sum in Eq. 18 is a good approximation to a Gaussian beam traveling at an angle θ whose waist is located at (x_0, y_0). The total E-field at any point in the system is given by

$$E_{\text{tot}}(x,z) = \begin{cases} \text{Real}\{\sum \Delta k_x \left(E_1^R(k_x)\exp[i(k_x x + k_{1z}^R z)] + E_1^L(k_x)\exp[i(k_x x + k_{1z}^L z)]\right)\}, & z \le 0 \\ \text{Real}\{\sum \Delta k_x \left(E_2^R(k_x)\exp[i(k_x x + k_{2z}^R z)] + E_2^L(k_x)\exp[i(k_x x + k_{2z}^L z)]\right)\}, & 0 \le z \le d \\ \text{Real}\{\sum \Delta k_x E_3^R(k_x)\exp[i(k_x x + k_{3z}^R z)]\}, & z \ge d \end{cases} \tag{23}$$

where E_1^L, E_2^R, E_2^L, and E_3^R are calculated element-wise from $E_1^R(k_x)$ according to Eqs. 7-10. The beam plots in Fig. 2 of the main text are calculated pixel-by-pixel from the sum in Eq. 23, with the values of x and z indicating the location of the pixel. The resultant field is normalized to the maximum field value in the image, and displayed in color. The pulse video is calculated similarly, except that the field is Gaussian in space and time, and so the field must be sampled in both the spatial and temporal frequency domains. The calculation time is significantly longer for the pulse compared to the beam, and the simulations are only practical to run on a supercomputer.

The finite nature of the sampling has consequences which must be considered in order to be sure that our results are not affected by numerical artifacts. Firstly, the truncation of the Gaussian beam in k-space to the sampling width w_s leads to a convolution with a sinc function in the spatial domain. Therefore, the side-tail of our beam is not truly Gaussian; rather, the envelope of the side-tail is Gaussian but the side-tail itself exhibits periodic sinc-like fluctuations in intensity (which cannot be seen in Fig. 2 of the main text, but can be seen in logarithmic plots which resolve the small intensities of the side-tail). The sampling width chosen for Fig. 2 was $w_s = 2w_k$ (with $N_s = 501$). We made sure that other choices of the sampling width, $w_s = 3w_k$ and $4w_k$ (with proportionally larger N_s so that Δk_x remained constant), did not affect the behavior of the plots. Therefore, our conclusions are not affected by the precise value of the sampling width w_s. Secondly, the finite number of samples N_s

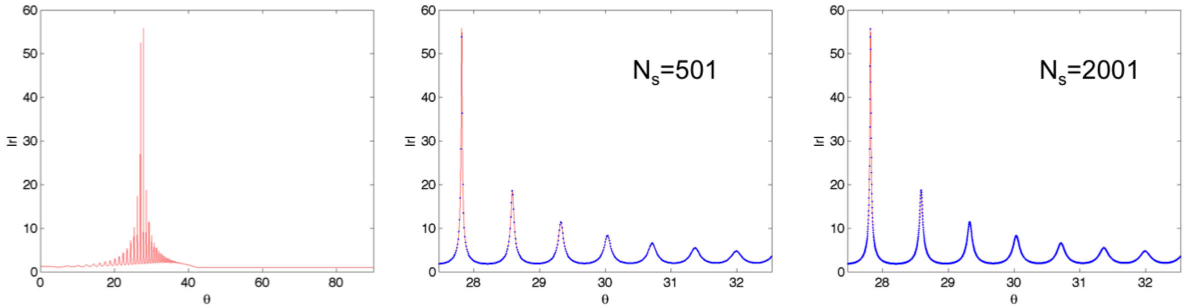

FIG. S3. A pictorial depiction of the Fourier domain sampling used to generate Fig. 2(b) of the main text. (a) A plot of $|r|$ vs. incidence angle θ for the case of $\lambda_o = 1$ μm, ϵ_1=ϵ_3=2.25, $\epsilon_2 = 1 - 0.01i$, and $d = 28$ μm. For a beam-width FWHM of 13.3 μm, the sampling width $w_s = 2w_k$ that we chose restricted the range of incidence angles of the planewaves used to construct the beam to $27.47° < \theta < 32.53°$. In plots (b) and (c), the blue dots overlaying the reflectivity plot indicate the incidence angles of the planewaves that were summed over for (b) $N_s = 501$ samples and (c) $N_s = 2001$ samples. The plot in Fig. 2(b) of the main text looks visually identical for both choices of sampling values.

implies the spectrum of k_x values is discrete, so the incident beam is periodic in space. This means that in the plots of Fig. 2 in the main text, there is not just one incident beam but an infinite number of them impinging on the slab, spaced periodically along the x-axis by a distance $2\pi/\Delta k_x = 2830$ μm. If the sampling is increased from $N_s = 501$ to 2001 while keeping $w_s = 2w_k$ constant (see Fig. S3), the distance between adjacent beams increases to 11330 μm, but the plots in both Figs. 2(a) and 2(b) of the main text look identical to the ones with 501 samples. Therefore, 501 samples is sufficient in this case to ensure that the (periodically repeated) beams do not interfere with each other, and that the plot is a good representation of the field of a single beam.

---

[1] J.B. Pendry. "Negative refraction makes a perfect lens." Phys. Rev. Lett. **85**, 3966-69 (2000).

[2] D.R. Smith, D. Schuring, M. Rosenbluth, S. Schultz, S.A. Ramakrishna, and J.B. Pendry. "Limitations on subdiffraction imaging with a negative refractive index slab." Appl. Phys. Lett. **82**, 1506-1508 (2003).

[3] N. Garcia and M. Nieto-Vesperinas. "Left-Handed Materials Do Not Make a Perfect Lens." Phys. Rev Lett. **88**, 20, 207403 (2002).

[4] X. Jiang, Q. Li and C.M. Soukoulis. "Symmetry between absorption and amplification in disordered media." Phys. Rev. B **19**, 9007 (1999).

[5] S.A. Ramakrishna. "On the dual symmetry between absorbing and amplifying random media." Pramana **62**, 1273-1279, (2004).

[6] J.C.J. Paasschens, T. Sh. Misirpashaev and C.W.J. Beenakker. "Localization of light: dual symmetry between absorption and amplification." Phys. Rev. B **54**, 11887 (1996).

[7] P. Pradhan and N. Kumar. "Localization of light in coherently amplifying random media." Phys. Rev. B **50**, 9644 (1994).

[8] Z.Q. Zhang. "Light amplification and localization in randomly layered media with gain." Phys. Rev. B **52**, 7960 (1995).

[9] Y. Chen, P. Fischer, and F.W. Wise. "Negative Refraction at Optical Frequencies in Nonmagnetic Two-Component Molecular Media." Phys. Rev. Lett. **95**, (2005).

[10] S.A. Ramakrishna. "Comment on 'Negative Refraction at Optical Frequencies in Nonmagnetic Two-Component Molecular Media.'" Phys. Rev. Lett. **98**, (2007).

[11] S.A. Ramakrishna and O.J.F. Martin. "Resolving the wave vector in negative index media." Opt. Lett. **30**, 19, 2626-2628 (2005).

[12] J.B. Geddes III, T.G. Mackay, and A. Lakhtakia. "On the refractive index for a nonmagnetic two-component medium: Resolution of a controversy." Opt. Comm. **280**, 120-125 (2007).

[13] B. Nistad and J. Skaar. "Causality and elecromagnetic properties of active media." Phys. Rev. E **78**, 036603 (2008).

[14] C.J. Koester. "Laser action by enhanced total internal reflection." IEEE J. Quantum Electron. **QE-2**, 580-4 (1966)

[15] P.R. Callary and C.K. Carniglia. "Internal reflection from an amplifying layer." J. Opt. Sci. Am. **66**, 8 (1976).

[16] J. Fan, A. Dogariu, and L.J. Wang. "Amplified total internal reflection." Opt. Express **11**, 299-308 (2003).

[17] K.J. Willis, J.B. Schneider, and S.C. Hagness. "Amplified total internal reflection: theory, analysis, and demonstration of existence via FDTD." Opt. Express **16**, 1903-14 (2008).

[18] A. Siegman. "Fresnel Reflection and Evanescent Gain." OPN, 38–45 (Jan. 2010).

[19] J.O. Grepstad and J. Skaar. "Total internal reflection and evanescent gain." Opt. Express **19**, 21404 (2011).

[20] E.P. Ippen and C.V. Shank. "Evanescent-field-pumped dye laser." Appl. Phys. Lett. **21**, 301 (1972).

[21] H.H. Choi, H.J. Kim, J. Noh and W. Jhe. "Amplified total internal reflection from a plane dielectric boundary." Jour. Korean Phys. Soc. **42**, 80 (2003).

Third Report of Referee A, received October 25th, 2013:

In their reply the authors clarify the main intend of their work. The argument goes as follows: we know that Fresnel formalism is applicable to some problems in optics. Let's take a problem where Fresnel formalism is likely not applicable, and let's see what happens then…

I agree that the question of applicability of Fresnel formalism is an important one. However, the current manuscript does not provide an answer; rather, it just starts the discussion. At this stage, I would suggest accepting this work to PRB brief reports section. However, in my opinion, to qualify for PRL, the authors need to provide the answer to the question of what will happen in real system with gain saturation, dispersion, and spontaneous emission present. I agree that this is a difficult problem to approach, but it does not mean that it cannot be done. Only after inclusion of these phenomena can one claim that the particular solution of Maxwell equations is a physical one. I still believe that the solution presented by the authors is not realizable in physical systems; I also believe that mathematically realizable solution is the one that I described in my previous report (I do agree that it represents a lossy material, and requires a rather sophisticated source); my claim is that (i) lossy materials exist and are stable in time, and (ii) if one manages to build this required source of excitation, he/she will be able to realize the results of calculations in the lab; on the other hand, my claim is that (iii) materials with net round-trip gain are unstable in time, and thus one cannot realize the scenario suggested by the authors in the lab; however, I do agree that these are just opinions and they must be verified. I am willing to take an FDTD solution as verification instead of real experiment (I would take experiment for verification), and would immediately recommend this manuscript for publication should any such verification prove the "refraction cancellation" scenario suggested by the movies.

In addition, with any revision I suggest that the authors revise their statement regarding amplified total internal reflection. This particular problem has convergent solution within Fresnel formalism, if one takes the correct cut in the complex plane when calculating kz. In particular, for nonmagnetic materials, the cut should go along negative imaginary axis of kz^2 plane; this way, evanescent waves are not allowed to exponentially grow inside the system, and the condition nu>1 is never fulfilled. This cut, proposed in APL v.91, 191103, has been verified by rigorous FDTD calculations in Opt.Exp. v.16, p.1903, and is also consistent with all previous experimental considerations. For example, in Refs.[14, 20], the pumping seems to be evanescent, so that the gain region is limited to the immediate proximity of the waveguide and there is no net round-trip gain (again, consistent with the proposed complex plane cut above).

Second Report of Referee B, received October 25th, 2013:

I have read the replies of the authors and the revised paper. I believe the authors have tried hard to address most of the comments as best they can and that the work deserves consideration for publication in PRL. The main issue seems to be whether the solution they present is physical or not (which I am sure the authors will agree is an important question to address), and the answer to this question is still under debate. In this regard, you may want to consult with another referee. As things stand at present, my advice is to consider publication as is.

Response of authors, sent October 29th, 2013:

Dear Editor and Referees,

We are happy that Referee B has agreed that our manuscript is suitable for publication in PRL, and that Referee A agrees that our results are correct and warrant publication in PRB. We hope, however, that we can convince Referee A that our results are of sufficient general interest to warrant publication in PRL.

Referee A attempts to summarize our approach in this manuscript as the following: "Lets take a problem where Fresnel formalism is likely not applicable, and lets see what happens then..." Of course, the only reason that it seems obvious to the referee that the Fresnel formalism is likely not applicable is because we have clearly identified the culprit–namely, a cavity with net roundtrip gain–that has caused peculiar results to surface in the literature. The referee could not know this, but we originally intended to investigate the problem of amplified TIR and had no idea about the role of the roundtrip coefficient. Once we discovered that amplified TIR is intrinsically linked to a cavity with $|\nu| > 1$, we set out to explore more generally the behavior of such a cavity. Most importantly, we found that this problem plays a significant but under-appreciated role in the literature with regards to Pendry's lens, negative refraction in nonmagnetic media, and the duality between loss and gain, in addition of course to amplified TIR.

We wholeheartedly agree with Referee A that the problem of what will happen in a real gain medium with gain saturation, dispersion, and spontaneous emission is an important and interesting problem. However, even if we managed to solve this problem, it would yield no additional insight to the problems that we have already solved and addressed in this manuscript. Our current results stand on their own and provide fundamentally new insights to problems that have been repeatedly discussed in the literature. Furthermore, this manuscript has already succeeded in generating a fruitful exchange between the referee and us. We ask the referee to consider the history of this manuscript: on your first revision, you rejected our results outright because the simulations were so counterintuitive that you felt they must be the result of Fourier artifacts. Now you agree that this is not the case, but do you not feel that the rest of the scientific community is deserving of seeing the incredibly counterintuitive predictions the Fresnel formalism can make? You agree that we have started the important discussion of when the Fresnel formalism is applicable. Even if we cannot answer this question definitively, aren't our results already relevant to the literature results that have completely ignored whether the Fresnel formalism is applicable to the problems they have attempted to solve? Might our results not serve as an important caution to physicists planning to use the Fresnel formalism in the future? We feel that the manuscript is worth publishing in PRL for these reasons alone, but we will also address the specific complaints of Referee A.

Referee A: "...my claim is that (iii) materials with net round-trip gain are unstable in time, and thus one cannot realize the scenario suggested by the authors in the lab; however, I do agree that these are just opinions and they must be verified."

Response: In our last revision, we provided in the supplementary information the example of a laser with destructive external feedback. This is a cavity with net roundtrip gain that is perfectly stable in time and has been realized in many labs! The referee did not comment on this example. While there are certainly scenarios in which a net roundtrip gain is unstable (such as a laser without feedback during the transient photon build-up time), the fact is that such situations are *not* universally unstable.

Referee A: "In addition, with any revision I suggest that the authors revise their statement regarding amplified total internal reflection. This particular problem has convergent solution within Fresnel formalism, if one takes the correct cut in the complex plane when calculating k_z. In particular, for nonmagnetic materials, the cut should go along negative imaginary axis of k_z^2 plane; this way, evanescent waves are not allowed to exponentially grow inside the system, and the condition $|\nu| > 1$ is never fulfilled. This cut, proposed in APL v. 91, 191103, has been verified by rigorous FDTD calculations in Opt. Exp. v. 16, p. 1903, and is also consistent with all previous experimental considerations. For example, in Refs. [14, 20], the pumping seems to be evanescent, so that the gain region is limited to the immediate proximity of the waveguide and there is no net round-trip gain (again, consistent with the proposed complex plane cut above)."

Response: In our manuscript, we treat the problem of TIR from an amplifying slab, not of a semi-infinite medium. In the slab problem, no branch cut is necessary because both counter-propagating wavevectors exist. We have shown that such a slab will have net roundtrip gain when excited by light incident above the critical angle, and we explain mathematically why k_{2z}^L dominates the behavior of the slab when $|\nu| \gg 1$. Now, the referee thinks that the Fresnel formalism does not apply to this slab, but that the Fresnel formalism *can* be used to treat the single-surface problem, for which he claims that k_{2z}^L is the correct choice for the transmitted wavevector. Let's put it this way: for a 10-meter thick slab (or some other absurdly large thickness) we claim that k_{2z}^L will be the only relevant wavevector as a result of the pre-excitation, but the referee thinks that our solution is likely not physical. If the slab is infinitely thick instead of "only" 10 meters thick, then the referee believes that k_{2z}^L is the transmitted wavevector, and moreover that this solution is perfectly physical. For one, we feel that nobody can claim to have an adequate understanding of the single-surface problem unless they can also explain the slab problem. Secondly, the references the referee cites are not as rigorous as he believes, and we will spend the remainder of our response addressing this point. We have intentionally chosen not to include this information in our supplementary information because we want to include it in a separate paper devoted specifically to amplified TIR.

Alternative explanation for FDTD and experimental results on amplified TIR

The controversy over single-surface amplified total internal reflection (TIR) concerns the following question: when light in a transparent, high-index medium (medium one) is incident above the critical angle on a semi-infinite lower-index amplifying medium (medium two), is the correct choice for the transmitted wavevector k_{2z}^R–resulting in an exponentially growing wave away from the interface in medium two–or k_{2z}^L, resulting in an exponentially decaying wave in medium two along with an amplified reflection in medium one? Below the critical angle it is universally agreed that the correct choice is k_{2z}^R, but above the critical angle many have argued in favor of k_{2z}^L in order to avoid an exponentially-growing evanescent wave in medium two.

In our manuscript, we briefly treat the problem of TIR from an amplifying slab, which is of course different than an amplifying semi-infinite medium but nevertheless relevant. Within the slab both wavevectors k_{2z}^R and k_{2z}^L exist, but the wave with wavevector k_{2z}^R has a vanishingly small amplitude. Furthermore, in the limit as the slab thickness $d \to \infty$, the amplitude of the right-propagating wave goes to zero and only the wavevector k_{2z}^L remains. (This statement is not controversial; it is a mathematical fact.) Some people have concluded from this fact that the correct choice for the wavevector in the single-surface problem (i.e., when medium two is truly semi-infinite) must be k_{2z}^L [1]. We, on the other hand, argue that the solution obtained by taking the limit $d \to \infty$ is *not* equivalent to the solution of the single-surface problem. Instead, we argue that the pre-excitation mechanism that we demonstrated below the critical angle should be equally applicable above the critical angle; if one accepts this premise, then the vanishingly small amplitude of the right-propagating wave and dominance of the left-propagating wave is the direct result of the multiple reflections of the pre-excited field. (We emphasize that whether or not the pre-excited field turns out to be a physical solution–which must be determined by experiment–it remains the direct mathematical consequence of the Fresnel solution to this problem. As such, it merits serious consideration in any discussion of amplified TIR.) In a semi-infinite medium this pre-excitation mechanism cannot occur; our argument therefore suggests that k_{2z}^R is the transmitted wavevector in the single-surface problem.

Most theoretical investigations of the amplified TIR problem apply the Fresnel formalism, which is sensible because the controversy itself is a direct consequence of the ambiguity permitted by the Fresnel formalism in the sign of the transmitted wavevector. However, finite difference time domain (FDTD) methods should in principle be capable of resolving the controversy as well, and one such investigation by Willis et al [2] concluded that the reflection from a semi-infinite gain medium is amplified and therefore k_{2z}^L is the transmitted wavevector. In the FDTD simulation, the gain medium is terminated by a perfectly matched layer (PML) which, it is argued, is practically equivalent to a semi-infinite gain medium. In most FDTD simulations, we agree that the residual reflectivity of the PML is small enough to justify this practical equivalence. In the peculiar case of TIR from an amplifying medium, however, we will demonstrate that even a very small reflection from the PML will substantially affect the results.

We will use the same parameters as those used in [2]: $\lambda_o = 900.28$ nm (equivalent to $f_o = 3.33 \times 14$ Hz), $\epsilon_1 = 4$, $\epsilon_2' = 2$, and $\epsilon_2'' = -0.027i$ (equivalent to a conductivity $\sigma_2 = -500$

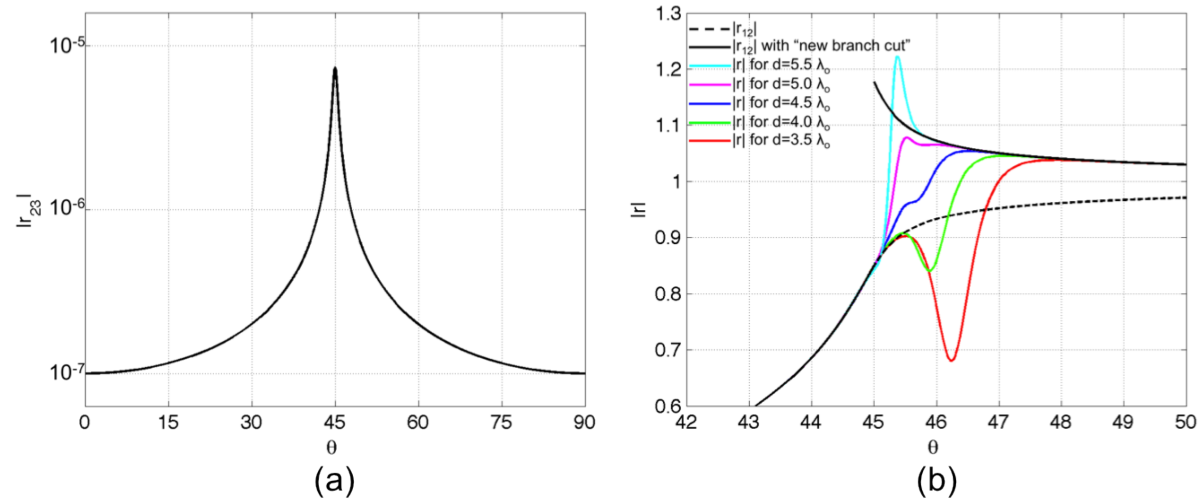

FIG. S1. The material parameters are $\epsilon_1 = 4$, $\epsilon_2 = 2 - 0.027i$, and $\epsilon_3 = 1.9999992000 - 0.0269999892i$, where material three is chosen to simulate an imperfect PML, and $\lambda_o = 902.28$ nm. (a) The reflectivity $|r_{23}|$, which is the residual reflectivity of the "PML" in our model, plotted against the incidence angle θ in medium one for s-polarization. (b) The reflectivity of the incident planewave in medium one $|r|$ vs. θ near $\theta_c = 45°$ for various thicknesses d of medium two. Also shown are the two candidate curves for the reflectivity r_{12} in the single-surface problem.

S/m at frequency f_o). Our goal is to demonstrate the potentially pernicious effects of the imperfect PML, not by doing our own FDTD simulation, but rather by solving a nearly equivalent scenario using the Fresnel approach to the three-layer problem. To model the effect of the PML, we introduce a third medium that is nearly–but not exactly–impedance-matched to medium two. With $\epsilon_3 = 1.9999992000 - 0.0269999892i$, the reflection coefficient r_{23} at normal incidence is 10^{-7}. At other angles of incidence θ–where θ is the incidence angle of the wave arriving at the 1-2 interface, not the 2-3 interface–the reflection coefficient r_{23} is plotted in Fig. S1(a). With this particular model we have chosen for our "PML," we see that r_{23} increases with θ to a maximum just below 10^{-5} at the critical angle $\theta_c = 45°$, then decreases again with increasing θ. In fact, the reflectivity of the PML used in FDTD simulations usually exhibits a monotonic increase with incidence angle; all that matters for our purposes, however, is that the reflectivity r_{23} is quite small at all angles.

The reflection coefficient of a planewave in medium one incident upon this structure is given by the usual three-layer formula

$$r = \frac{r_{12} + r_{23}\exp(2ik_{2z}^R d)}{1-\nu} \tag{1}$$

where

$$\nu = r_{21}r_{23}\exp(2ik_{2z}^R d). \tag{2}$$

In Fig. S1(b) we have plotted the reflection amplitude $|r|$ for an s-polarized incident planewave as a function of incidence angle, for various slab thicknesses d between $3.5\lambda_o$ and $5.5\lambda_o$. For comparison, we have also plotted the single-surface reflection coefficient r_{12} (in other words, the analytical result for the case where medium two is semi-infinite) for the two competing theories: when k_{2z}^R is the transmitted wavevector at all angles (dashed black), and when k_{2z}^L is the transmitted wavevector above the critical angle (solid black), which we will call the "new branch cut" theory (borrowing the terminology from [2]). Regardless of the thickness d, all the curves approach the solid black line at large enough incidence angle. This is because at large angles the imaginary part of k_{2z}^R is large and negative, causing the roundtrip propagation gain $\exp(2ik_{2z}^R d)$ to be enormous; more specifically, it is large enough that the net roundtrip gain $|\nu|$ can exceed one despite the fact that r_{23} is on the order of 10^{-6}. As we have shown in the manuscript, when $|\nu| \gg 1$, the dominant wavevector in the slab is k_{2z}^L and r approaches the "new branch cut" solution. We emphasize that *the amplified reflection can be entirely explained by the three-layer picture, even when $|r_{23}|$ is as small as the reflection from a typical PML layer in an FDTD simulation.*

Because the three-layer theory and the "new branch cut" theory predict the same reflectivity at large incidence angles (as can be seen in Fig. S1(b) for $\theta > 47°$), to accurately discriminate between the two theories one needs to look at the reflectivity close to the critical angle, roughly between $45° < \theta < 47°$. We have reproduced the FDTD simulation results presented in Fig. 8(b) of [2] in our Fig. S2. The red circles are FDTD simulation results for the reflectivity of a Gaussian beam, and the theoretical red dashed line is a weighted sum of the reflectivities obtained by applying the "new branch cut" theory to the constituent planewaves of the Gaussian beam. We feel that it is important to point out that the FDTD data points which deviate the farthest from the theoretical curve are those at $\theta = 45°$ and $46°$, which are the most important FDTD data points needed to accurately discriminate between the three-layer theory and single-surface "new branch cut" theory. Alternatively, we propose that the three-layer curves for $d = 5.5\lambda_o$ and $d = 5\lambda_o$ in Fig. S1(b) qualitatively

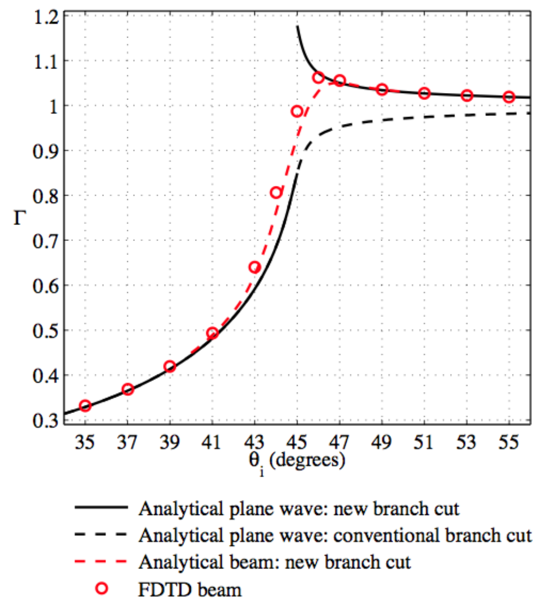

FIG. S2. Reproduction of Fig. 8(b) in [2]. Γ is the reflection amplitude that we have been calling $|r|$. The red circles are FDTD simulation results for a Gaussian beam. Because the beam comprises many planewaves with a spread of incidence angles, the theoretical curve (red dashed line) is obtained by applying the "new branch cut" theory to each constituent planewave and summing the contributions.

resemble the behavior of the FDTD simulation, and may in fact prove to be a better quantitative fit to the FDTD results once the true reflectivity of the PML is used in the three-layer calculation, and the reflectivity is averaged over all constituent planewaves of the Gaussian beam. (At this time we cannot do this ourselves because we do not know the exact thickness d used in the FDTD simulation, nor the PML reflectivity curve vs. θ.)

It is therefore not implausible that the FDTD simulation simply yields the results expected from the analytical solution of the three-layer problem. In other words, the amplified reflection observed by Willis could be explained by the small but non-zero reflectivity of the PML, without the need to invoke any theory on single-surface amplified TIR. Simply put, the reflectivity of the PML layer is not small enough for the FDTD simulation to accurately represent the physics of the single-surface problem. Before accepting Willis' conclusions, a more rigorous study of the residual reflectivity of the PML at incidence angles near the critical angle is needed. It may also be helpful to see how the simulation results are affected by varying the thickness d of the slab. In Fig. S1(b), we included the reflectivity curves for $d = 4\lambda_o$ and $d = 3.5\lambda_o$ because they show a curious relative minimum shortly after the critical angle. Perhaps this effect can be seen in the FDTD simulations, but of course it may depend sensitively on the true reflection coefficient of the PML as a function of θ.

We mention that the three-layer analysis we have presented here can also explain many experimental results [3, 4] again without recourse to any theory of single-surface reflection. In an experiment the gain medium must be pumped, and there will inevitably be a variation in the pumping power with depth in medium two. Therefore, the gain will vary spatially. Even if this variation seems negligible, it will be enough to generate a small reflection r_{23}, and then the results of our analysis apply. Since it is likely that no experiment will ever be capable of creating a gain medium homogeneous enough to prevent the reflection r_{23}, we expect all experiments to yield amplification of the totally internally reflected beam. However, we believe that this is a result of the multilayer reflection, and not a property of the single-surface reflection.

We hope that these arguments have cast a reasonable doubt in your mind on the existing explanations of FDTD and experimental results on amplified TIR, and furthermore reinforced the predictive power and importance of the roundtrip coefficient in these problems. Thank you again for your continued consideration of our manuscript.

Sincerely,

Tobias S. Mansuripur, Graduate Student

Department of Physics

Harvard University

---

[1] P.R. Callary and C.K. Carniglia. "Internal reflection from an amplifying layer." J. Opt. Sci. Am. **66**, 8 (1976).

[2] K.J. Willis, J.B. Schneider, and S.C. Hagness. "Amplified total internal reflection: theory, analysis, and demonstration of existence via FDTD." Opt. Express **16**, 1903-14 (2008).

[3] C.J. Koester. "Laser action by enhanced total internal reflection." IEEE J. Quantum Electron.

QE-2, 580-4 (1966)

[4] H.H. Choi, H.J. Kim, J. Noh and W. Jhe. "Amplified total internal reflection from a plane dielectric boundary." Jour. Korean Phys. Soc. **42**, 80 (2003).

Fresnel reflection from a cavity with net roundtrip gain

Tobias S. Mansuripur[1,*] and Masud Mansuripur[2]

[1] Department of Physics, Harvard University, Cambridge, MA 02139
[2] College of Optical Sciences, The University of Arizona, Tucson, AZ 85721

A planewave incident on an active etalon with net roundtrip gain may be expected to diverge in field amplitude, yet applying the Fresnel formalism to Maxwell's equations admits a convergent solution. We describe this solution mathematically, and provide additional insight by demonstrating the response of such a cavity to an incident beam of light. Cavities with net roundtrip gain have often been overlooked in the literature, and a clear understanding of their behavior yields insight to amplified total internal reflection, negative refraction in nonmagnetic media, and negative-index lenses.

The Fresnel coefficients govern the reflection and transmission of light for the simplest possible scenarios: at planar interfaces between homogeneous media. Despite this simplicity, some interesting solutions have been discovered only recently, such as 1) the amplification of evanescent waves in a passive, negative-index slab [1–3] and 2) a duality between loss and gain leading to the localization of light in both cases [4–8]. In addition, controversy regarding the proper choice of the wavevector in active media has persisted in relation to the possibility of 3) negative refraction in nonmagnetic media [9–13] as well as 4) single-surface amplified total internal reflection (TIR) [14–19]. It turns out that all four of these cases share a common thread: the presence of a cavity whose roundtrip gain exceeds the loss. In this Letter, we explore more generally the response of such a cavity to an incident beam of light.

To begin, we establish a convention that allows us to more clearly discuss the direction of energy flow. We associate with each wavevector a superscript R (L) to indicate that the wave carries energy to the right (left), namely, that for which the time-averaged z-component of the Poynting vector is positive (negative). (See supplementary information for details.) For the single-surface problem, shown in Fig. 1(a), the incident wavevector in medium one is $\boldsymbol{k}_1^R = k_x\hat{\boldsymbol{x}} + k_{1z}^R\hat{\boldsymbol{z}}$, and the reflected wavevector is $\boldsymbol{k}_1^L = k_x\hat{\boldsymbol{x}} + k_{1z}^L\hat{\boldsymbol{z}}$, where $k_{1z}^L = -k_{1z}^R$. The real-valued component k_x, once determined by the incident wave, is the same for all wavevectors in the system. For the transmitted wavevector, the dispersion relation offers two choices for k_{2z},

$$k_{2z} = \pm\sqrt{(\omega/c)^2\mu_2\epsilon_2 - k_x^2}, \quad (1)$$

where ω is the angular frequency, c is the speed of light in vacuum, and μ_2 and ϵ_2 are the relative permeability and permittivity. It is universally agreed that the correct choice for k_{2z} in the single-surface problem is k_{2z}^R (i.e., that the transmitted energy flows away from the interface), irrespective of the material parameters or the nature of the incident wave, except possibly in the case of amplified TIR, for which there remains debate. Due to this controversy, let us postulate for now that k_{2z}^R is the correct choice in all cases, so that we can unambiguously define the single-surface Fresnel reflection and transmission coefficients

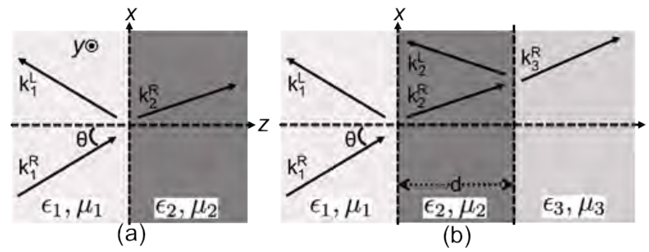

FIG. 1. Geometry of the (a) single-surface and (b) cavity problems. All media are infinite in the x and y-directions. The arrows denote the wavevectors of the planewaves present in each layer.

$$r_{\ell m} = \frac{\tilde{k}_{\ell z}^R - \tilde{k}_{mz}^R}{\tilde{k}_{\ell z}^R + \tilde{k}_{mz}^R}, \; t_{\ell m} = \frac{2\tilde{k}_{\ell z}^R}{\tilde{k}_{\ell z}^R + \tilde{k}_{mz}^R} \quad (2)$$

where we have generalized the result for incidence medium ℓ and transmission medium m. For s-polarization we have defined $\tilde{k}_{nz} \equiv k_{nz}/\mu_n$, while for p-polarization $\tilde{k}_{nz} \equiv k_{nz}/\epsilon_n$.

We now consider the case of light incident on a cavity, shown in Fig. 1(b). The total E-field resulting from an s-polarized incident wave in medium one with amplitude E_1^R is given by

$$E_y(x,z) = \begin{cases} E_1^R \exp(ik_x x + ik_{1z}^R z) \\ \quad +E_1^L \exp(ik_x x + ik_{1z}^L z) & : z \le 0 \\ E_2^R \exp(ik_x x + ik_{2z}^R z) \\ \quad +E_2^L \exp(ik_x x + ik_{2z}^L z) & : 0 \le z \le d \\ E_3^R \exp[ik_x x + ik_{3z}^R (z-d)] & : z \ge d \end{cases} \quad (3)$$

where the time-dependence factor $\exp(-i\omega t)$ has been omitted. The most direct route to solve for the four unknown wave amplitudes is to enforce Maxwell's boundary conditions at $z = 0$ and $z = d$, which yields four equations that can be solved for the four unknowns. The resulting reflection coefficient from the slab can be ex-

* mansuripur@physics.harvard.edu

pressed in terms of the single-surface Fresnel coefficients as

$$r \equiv \frac{E_1^L}{E_1^R} = \frac{r_{12} + r_{23}\exp(2ik_{2z}^R d)}{1-\nu} \quad (4)$$

where

$$\nu = r_{21}r_{23}\exp(2ik_{2z}^R d) \quad (5)$$

is referred to as the roundtrip coefficient; the amplitude of a planewave circulating in the slab is multiplied by this factor after each roundtrip in the absence of any sources outside the slab. (Although we explicitly discuss s-polarized light, our conclusions as well as Eqs. 4 and 5 hold for both polarization states.) We emphasize that the reflection coefficient given by Eq. 4 is a valid solution to Maxwell's equations for any value of ν. The roundtrip coefficient ν has an important physical meaning, and intuitively one would expect three different regimes of behavior when the magnitude of ν is less than, equal to, or greater than one. The case where $|\nu| < 1$ governs passive slabs (in most but not all cases) and sufficiently weakly amplifying slabs. When $\nu = 1$ the slab behaves as a laser and emits light even in the absence of an incident wave, which manifests itself mathematically as an infinitely large reflection amplitude. The case where $|\nu| > 1$, however, has received scant attention in the literature [4, 15].

Perhaps the reason for the neglect of the $|\nu| > 1$ steady-state solution is the seemingly intuitive assumption that there cannot be a steady-state solution when $|\nu| > 1$–due to gain saturation in a laser, for instance. (See supplementary information.) This assumption is only reinforced by examining a second well-known solution method for the reflection coefficient that decomposes the reflected wave amplitude E_1^L into a sum over partial waves, yielding the reflection coefficient

$$r = r_{12} + t_{12}t_{21}r_{23}\exp(2ik_{2z}^R d)\sum_{m=0}^{\infty}\nu^m. \quad (6)$$

Heuristically, the first term r_{12} (hereinafter referred to as the "specular" partial wave) of Eq. 6 results from the single-surface reflection of the incident wave at the 1-2 interface, and the geometric series accounts for the contributions to the reflected wave following multiple roundtrips within the slab. When $|\nu| < 1$, the geometric series in Eq. 6 converges to $(1-\nu)^{-1}$, giving the same result as found by matching the boundary conditions in Eq. 4. When $|\nu| > 1$, however, the geometric series diverges and the reflection coefficient is infinite. Intuitively, this divergence seems reasonable, since we expect any light that couples into a slab with $|\nu| > 1$ to be amplified after each roundtrip, and therefore grow without bound. Nevertheless, Eq. 4 yields a finite reflection coefficient even when $|\nu| > 1$, so how can we reconcile these two very different solutions?

In fact, the partial wave method *can* be used to find the $|\nu| > 1$ convergent solution. First, note that the reflection coefficient given by Eq. 4 is invariant under the transformation $k_{2z}^R \leftrightarrows k_{2z}^L$. (This is not surprising as it can be interpreted simply as a relabeling of the waves $E_2^R \leftrightarrows E_2^L$ in Eq. 3 that does not affect the final result.) Applying this same transformation to the partial wave sum of Eq. 6 [15], we can express the reflection coefficient as

$$r = r'_{12} + t'_{12}t'_{21}r'_{23}\exp(2ik_{2z}^L d)\sum_{m=0}^{\infty}\nu'^m, \quad (7)$$

where the prime indicates the substitution $k_{2z}^R \to k_{2z}^L$. Because the new roundtrip coefficient, $\nu' = r'_{21}r'_{23}\exp(2ik_{2z}^L d)$, is equal to ν^{-1}, in cases where $|\nu| > 1$ the primed partial wave sum of Eq. 7 will converge to the reflection coefficient of Eq. 4. (Incidentally, this duality between ν and ν^{-1} provides a simple mathematical explanation for the loss/gain duality observed by others [4–8]).

The physical implications of the substitution $k_{2z}^R \to k_{2z}^L$ in the partial wave sum can best be seen by examining the behavior of a "finite-diameter" beam of light incident obliquely on the slab. By numerically superposing a finite number of planewave solutions to Eq. 3 with appropriate amplitudes and incidence angles in the range $27.47° < \theta < 32.53°$ (see supplementary information), we create a Gaussian (to within the sampling accuracy) beam incident on the slab at 30° with a full-width at half-maximum beam-diameter of 13.3 μm. All media are nonmagnetic, and we choose $\epsilon_1 = \epsilon_3 = 2.25$ and the slab to be an amplifying medium with $\epsilon_2 = 1 - 0.01i$. The free-space wavelength of the beam is $\lambda_o = 1$ μm. We can examine the transition at $|\nu| = 1$ simply by varying d, since both $|r_{21}|$ and $|r_{23}|$ are less than one (and independent of d), whereas $|\exp(2ik_{2z}^R d)|$ (and hence ν) increases monotonically with d (because k_{2z}^R has a negative imaginary part). A plot of the field $E_y(x,z)$ at one instant of time is shown in Fig. 2(a) for $d = 19$ μm, which was chosen so that $|\nu|$ is slightly less than one for all constituent planewaves of the beam ($0.46 < |\nu| < 0.99$). The arrows overlying the plot point in the direction of the time-averaged Poynting vector within their vicinity, indicating the direction of energy flow in the system, and the incident beam is uniquely identified by the white arrow. The beam behaves as we expect it to: the incident beam strikes the slab near ($x = 0$, $z = 0$), giving rise to a specularly reflected beam as well as a refracted beam that 'zig-zags' up the slab, which in turn generates a reflected beam in medium one each time it strikes the 2-1 interface. (The field amplitude is plotted on a linear scale, and so the incident beam as well as the specularly reflected beam appear faint relative to the subsequently amplified portions of the beam.) Each of these reflected beams can intuitively be associated with a term of the partial wave expansion of Eq. 6–either the specular term or the mth term of the geometric series.

In Fig. 2(b) all parameters are kept the same except the slab thickness is increased to $d = 28$ μm, resulting in $|\nu| > 1$ for all constituent planewaves of the Gaussian

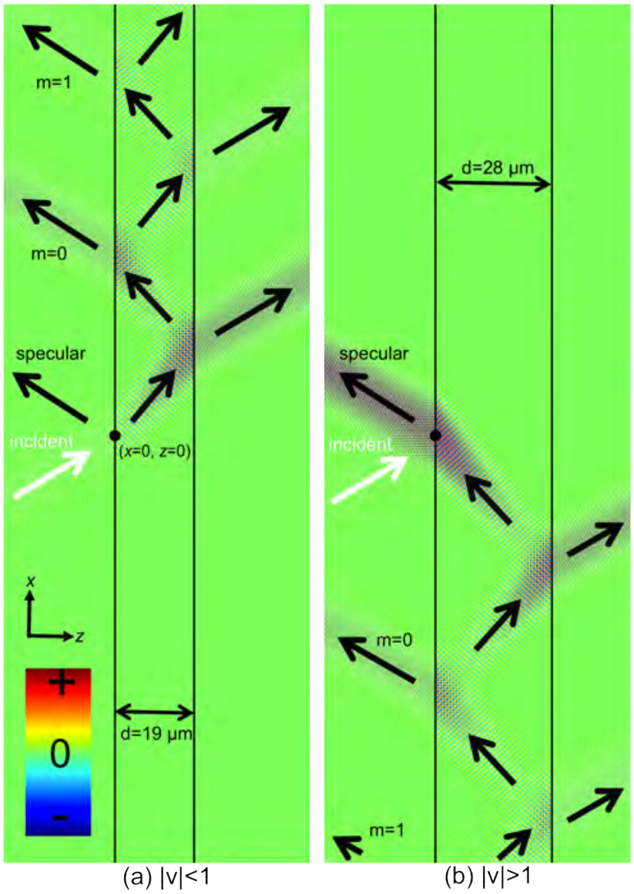

FIG. 2. Plots of the field $E_y(x, z)$ at one instant of time for a Gaussian beam (indicated with the white arrow) incident on an amplifying slab for which (a) $|\nu| < 1$ ($d = 19$ μm) and (b) $|\nu| > 1$ ($d = 28$ μm). The black dot indicates the origin of the coordinate system.

beam ($1.01 < |\nu| < 2.58$). Based solely on the plot of the field amplitude and not on the direction of energy flow indicated by the arrows, it may appear that the incident beam strikes the interface and negatively refracts in the slab, then zig-zags downwards in the $-\hat{\boldsymbol{x}}$ direction, giving rise to many reflected beams in medium one (and transmitted beams in medium three) which emanate from points on the slab with $x < 0$. Such an explanation was offered for simulations similar to ours [12, 13] to attempt to justify negative refraction in an active, nonmagnetic medium. However, by analyzing the Poynting vector we see that the energy in the beam zig-zags *up* the slab, so this phenomenon is distinct from negative refraction, despite the similarity in the positions of the reflected and transmitted beams. (In the supplementary information, a video of the time-dependent behavior of a "finite-duration" pulse of light more clearly illustrates that energy flows in the $+x$-direction.) We will refer to the field in the slab at $x < 0$ as the "pre-excitation," so-called because it occurs before the central lobe of the incident beam arrives at the slab. (We will discuss possible mechanisms for the origin of the pre-excitation shortly, but for now simply accept it as a necessary consequence of the Fresnel solution.) Each reflected beam in Fig. 2(b) can be associated either with the specular term r'_{12} or with the mth term of the primed partial wave expansion in Eq. 7. In particular, note that $r'_{12} = r_{12}^{-1}$; since $|r_{12}| < 1$ (in most cases of practical interest), this means that $|r'_{12}| > 1$, and so the primed partial wave expansion mathematically predicts the amplification of the specularly reflected beam only when $|\nu| > 1$. From Fig. 2(b) we see that this amplification occurs because the specular beam receives energy from the transmission of the pre-excited field through the 2-1 interface.

It is interesting to compare a lossy and an amplifying slab in the limit as $d \to \infty$. In a lossy slab (for which $\mathrm{Im}(k_{2z}^R) > 0$), the roundtrip coefficient $\nu \to 0$ as $d \to \infty$, and so the reflection coefficient r approaches the single-surface solution r_{12}, as expected, because the geometric series in Eq. 6 makes no contribution. In a gainy slab (for which $\mathrm{Im}(k_{2z}^R) < 0$), $\nu \to \infty$ as $d \to \infty$, but $\nu' \to 0$ and so we see from Eq. 7 that $r \to r'_{12}$ (which means that the field in the slab is dominated by the wavevector k_{2z}^L, i.e., $E_2^R/E_2^L \to 0$). The reason the limiting treatment of $d \to \infty$ for a gainy slab does not yield the proper single-surface reflection coefficient is that no matter how large one chooses to make d, the nonzero reflection r_{23} at the back-facet of the slab allows for the amplification of the pre-excited field; for $|\nu| \gg 1$, this results in the left-propagating wavevector k_{2z}^L dominating the behavior of the slab while the amplitude of the wave associated with k_{2z}^R diminishes substantially, and the reflection coefficient correspondingly approaches r'_{12}. Nevertheless, the right-propagating wave is essential in spite of its seemingly inconsequential amplitude, as it is responsible for generating the left-propagating wave by way of the back-facet reflection. In the case of two truly semi-infinite media (i.e., media one and two), the absence of a back-facet prevents any roundtrip amplification of the pre-excitation, so the only wavevector that exists in the transmission medium is k_{2z}^R–as agreed upon in all cases except possibly amplified TIR–and the single-surface reflection coefficient is correctly given by r_{12}, not r'_{12}.

The scenario under scrutiny in the "amplified TIR" problem concerns light in a high-index passive medium incident above the critical angle for TIR on a semi-infinite lower-index amplifying medium. Some have argued that the incident wave excites the wavevector k_{2z}^L in the transmission medium rather than the usual k_{2z}^R, resulting in the reflection coefficient r'_{21} and an accompanying amplified specular reflection [14–17, 19]. Before considering an amplifying medium of semi-infinite thickness, however, let us consider a finite-thickness amplifying slab and understand the role that ν plays in this problem. For the same parameters as those used in Fig. 2(b), $|\nu|$ increases monotonically with increasing incidence angle θ for s-polarized light; in particular, $|\nu|$ exceeds one as long as $\theta > 27.43°$. As θ approaches and surpasses the critical angle for TIR, $\theta_c = 41.8°$, $|\nu|$ quickly becomes extremely large due to the negatively increasing $\text{Im}(k_{2z}^R)$. (For $\theta = 41°$, $|\nu| = 9.34 \times 10^3$, and for $\theta = 42°$, $|\nu| = 1.40 \times 10^{15}$.) More generally, TIR from a gainy slab is well within the regime $|\nu| >> 1$ (for any reasonable thickness d). This is then comparable to the case of large d, for which we demonstrated in the previous paragraph that the existence of the left-propagating k_{2z}^L relies on the nonzero back-facet reflection r_{23} [18], resulting in a reflection coefficient approaching r'_{12}. When the amplifying medium is semi-infinite, however, $\nu = 0$ because r_{23} is effectively zero, therefore the pre-excitation does not occur and consequently the wave with wavevector k_{2z}^L does not exist. This suggests that k_{2z}^R is the correct choice for the transmitted wavevector even in the case of TIR from an amplifying medium.

The Fresnel solution for a slab with $|\nu| > 1$ is a steady-state harmonic solution, in the sense that if the field distribution presented in Fig. 2(b) exists at time t_o, then as time is evolved forward the field at each point in space will vary harmonically with frequency ω. We have not yet addressed the question of how the pre-excited field is established in the first place, beginning from a state with no electromagnetic fields anywhere in space. One possibility that does not violate causality is to consider that the Gaussian beam does not have a truly finite spatial width, but rather a rapidly decaying "side-tail" in the direction normal to the propagation direction. The side-tail is capable of injecting a small amount of energy into the slab at positions $x \ll 0$. When $|\nu| > 1$, light in the slab gains more during one roundtrip than it loses to transmission at both facets, and so this initially small amount of energy is amplified, resulting in the pre-excited field. For a beam or pulse at normal incidence, the leading edge (also known as the Sommerfeld front-runner) can perhaps play a similar role. The key point is that when $|\nu| > 1$, our intuition about the arrival time and arrival position of the beam (or pulse) misleads us because amplification by the slab acts on typically negligible field amplitudes to dramatically alter the character of the field. Finite-difference time-domain methods, applied to a slab whose dispersion obeys the Kramers-Kronig relations, are perhaps best suited to investigate the causal evolution of the pre-excitation. We note also that the Fresnel formalism, by implicitly beginning with the time-harmonic subset of Maxwell's equations, can only elucidate the non-divergent solutions to the full time-dependent equations. There can certainly exist divergent solutions, as demonstrated by finite-element simulations of a wave with a well-defined start-time incident normally on a slab with $|\nu| > 1$ [4]. Therefore, it is worth questioning the stability of the convergent solution: can it be seen in practice, or will it be obscured by other solutions? Some may also question whether a cavity with net roundtrip gain can exist in the steady-state in the first place; we note that the experimental work already done on the amplification of evanescent waves [14, 20, 21] (a regime for which $|\nu| > 1$) has not detected any instabilities, due to spontaneous emission or otherwise. A laser with destructive external self-feedback is also a stable system for which $|\nu| > 1$, which we discuss in the supplementary information. Finally, we emphasize that $|\nu|$ can exceed one even in passive media provided $|r_{21}r_{23}|$ exceeds one, which can happen for incident evanescent waves near surface plasmon resonances and therefore provides insight to the behavior of the negative-index lens [1–3]. This will be treated fully in future work, but we mention it here to prevent the reader from simply dismissing the $|\nu| > 1$ scenario as an unphysical mathematical peculiarity of gain media. In the end, only an experiment can determine whether the solution we have presented is a physical one; at the very least, we hope that our application of the Fresnel formalism to the unintuitive case of a cavity whose roundtrip gain exceeds the loss has offered fresh perspective on a few persistent controversies, and perhaps generated interesting ideas for future experiments.

The pulse simulation was run on the Odyssey cluster supported by the Harvard FAS Research Computing Group. TSM is supported by an NSF Graduate Research Fellowship. We thank Alexey Belyanin and Steven Byrnes for helpful discussions.

---

[1] J.B. Pendry. "Negative refraction makes a perfect lens." Phys. Rev. Lett. **85**, 3966-69 (2000).

[2] D.R. Smith, D. Schuring, M. Rosenbluth, S. Schultz, S.A. Ramakrishna, and J.B. Pendry. "Limitations on subdiffraction imaging with a negative refractive index slab." Appl. Phys. Lett. **82**, 1506-1508 (2003).

[3] N. Garcia and M. Nieto-Vesperinas. "Left-Handed Materials Do Not Make a Perfect Lens." Phys. Rev Lett. **88**, 20, 207403 (2002).

[4] X. Jiang, Q. Li and C.M. Soukoulis. "Symmetry between absorption and amplification in disordered media." Phys. Rev. B **19**, 9007 (1999).

[5] S.A. Ramakrishna. "On the dual symmetry between absorbing and amplifying random media." Pramana **62**, 1273-1279, (2004).

[6] J.C.J. Paasschens, T. Sh. Misirpashaev and C.W.J. Beenakker. "Localization of light: dual symmetry between absorption and amplification." Phys. Rev. B **54**, 11887 (1996).

[7] P. Pradhan and N. Kumar. "Localization of light in coherently amplifying random media." Phys. Rev. B **50**, 9644 (1994).

[8] Z.Q. Zhang. "Light amplification and localization in randomly layered media with gain." Phys. Rev. B **52**, 7960 (1995).

[9] Y. Chen, P. Fischer, and F.W. Wise. "Negative Refraction at Optical Frequencies in Nonmagnetic Two-Component Molecular Media." Phys. Rev. Lett. **95**, (2005).

[10] S.A. Ramakrishna. "Comment on 'Negative Refraction at Optical Frequencies in Nonmagnetic Two-Component Molecular Media.'" Phys. Rev. Lett. **98**, (2007).

[11] S.A. Ramakrishna and O.J.F. Martin. “Resolving the wave vector in negative index media.” Opt. Lett. **30**, 19, 2626-2628 (2005).

[12] J.B. Geddes III, T.G. Mackay, and A. Lakhtakia. “On the refractive index for a nonmagnetic two-component medium: Resolution of a controversy.” Opt. Comm. **280**, 120-125 (2007).

[13] B. Nistad and J. Skaar. “Causality and elecromagnetic properties of active media.” Phys. Rev. E **78**, 036603 (2008).

[14] C.J. Koester. “Laser action by enhanced total internal reflection.” IEEE J. Quantum Electron. **QE-2**, 580-4 (1966)

[15] P.R. Callary and C.K. Carniglia. “Internal reflection from an amplifying layer.” J. Opt. Sci. Am. **66**, 8 (1976).

[16] J. Fan, A. Dogariu, and L.J. Wang. “Amplified total internal reflection.” Opt. Express **11**, 299-308 (2003).

[17] K.J. Willis, J.B. Schneider, and S.C. Hagness. “Amplified total internal reflection: theory, analysis, and demonstration of existence via FDTD.” Opt. Express **16**, 1903-14 (2008).

[18] A. Siegman. “Fresnel Reflection and Evanescent Gain.” OPN, 38–45 (Jan. 2010).

[19] J.O. Grepstad and J. Skaar. “Total internal reflection and evanescent gain.” Opt. Express **19**, 21404 (2011).

[20] E.P. Ippen and C.V. Shank. “Evanescent-field-pumped dye laser.” Appl. Phys. Lett. **21**, 301 (1972).

[21] H.H. Choi, H.J. Kim, J. Noh and W. Jhe. “Amplified total internal reflection from a plane dielectric boundary.” Jour. Korean Phys. Soc. **42**, 80 (2003).

Fresnel reflection from a cavity with net roundtrip gain: Supplementary Information

Tobias S. Mansuripur[1,*] and Masud Mansuripur[2]

[1] *Department of Physics, Harvard University, Cambridge, MA 02139*

[2] *College of Optical Sciences, The University of Arizona, Tucson, AZ 85721*

[*] mansuripur@physics.harvard.edu

I. PRESCRIPTION FOR R AND L SUPERSCRIPTS

The energy flux of an s-polarized planewave whose E-field is given by

$$E_y(x,z,t) = E_0 \exp(ik_x x + ik_z z - i\omega t) \tag{1}$$

in a medium (ϵ, μ) is given by the time-average of the Poynting vector $\boldsymbol{S} \equiv \vec{E} \times \vec{H}$,

$$\langle \vec{S} \rangle = \frac{|E_0|^2}{2\omega\mu_0} e^{-2\mathrm{Im}(k_z)z} \left(\mathrm{Re}\left[\frac{k_x}{\mu}\right] \hat{x} + \mathrm{Re}\left[\frac{k_z}{\mu}\right] \hat{z} \right). \tag{2}$$

Therefore, energy flows in the $+z$-direction ('to the right,' in our convention) when $\mathrm{Re}(k_z/\mu) > 0$, and we denote the wavevector k_z which satisfies this condition with the superscript R. Any value k_z for which $\mathrm{Re}(k_z/\mu) < 0$ is accordingly labeled with a superscript L. It follows from these definitions that k_z^R must make an acute angle with μ in the complex plane. (For p-polarized light energy flows in the $+z$-direction when $\mathrm{Re}(k_z/\epsilon) > 0$, and so k_z^R makes an acute angle with ϵ in the complex plane.)

In cases where $\langle S_z \rangle = 0$, we must establish a prescription for resolving the ambiguity in the choice of superscript, which is best illustrated by an example. Consider the case where medium one is a lossless dielectric ($\epsilon_1 > 1$, $\mu_1 = 1$), medium two is vacuum, and the incident propagating wave satisfies $k_x > k_0$, where $k_0 \equiv \omega/c$, so that the two choices for k_{2z} are $\pm i\sqrt{k_x^2 - k_0^2}$. Both choices for k_{2z} yield pure evanescent waves and carry no energy along the z-direction. By adding a small amount of loss to medium two, so that $\epsilon_2 = 1 + i\epsilon_2''$ where $\epsilon_2'' > 0$, the two choices for k_{2z} deviate slightly from the imaginary axis as shown in Fig. S1(a). Now both waves carry non-zero energy along the z-direction; the first quadrant solution is k_{2z}^R (which can be seen quickly because it makes an acute angle with $\mu_2 = 1$) and the third quadrant solution is k_{2z}^L. Our prescription to establish k_{2z}^R for a true vacuum (i.e., $\epsilon_2'' = 0$) is to take the limit $\epsilon_2'' \to 0$, which yields k_{2z}^R as the solution along the positive imaginary axis.

Beware that if one adds a small amount of gain rather than loss to medium two, so that $\epsilon_2 = 1 + i\epsilon_2''$ where $\epsilon_2'' < 0$, then the two solutions for k_{2z} exist in the second and fourth quadrants as shown in Fig. S1(b), and in this case k_{2z}^R points predominantly along the *negative* imaginary axis. Thus, we see that the two limiting cases as gain or loss approaches zero do not yield the same result:

$$\lim_{\epsilon_2'' \to 0^+} k_{2z}^R = -\lim_{\epsilon_2'' \to 0^-} k_{2z}^R. \tag{3}$$

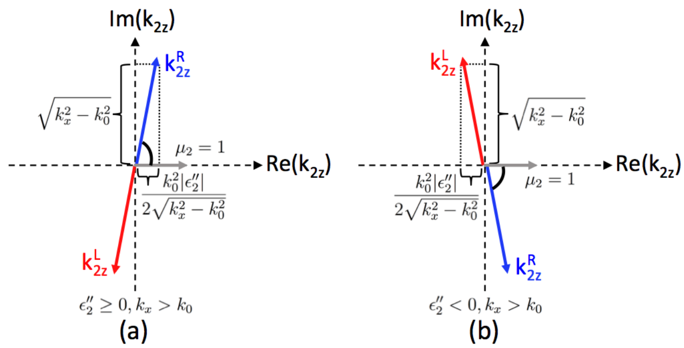

FIG. S1. Choosing the R or L label for an evanescent wave. (a) The two choices for k_{2z} are shown for the case of a slightly "lossy vacuum" ($\epsilon_2 = 1 + i\epsilon_2''$ where $\epsilon_2'' > 0$, $\mu_2 = 1$), for the case $k_x > k_0$. The first quadrant solution carries energy to the right and is labeled k_{2z}^R, and our prescription is to take the limit $\epsilon_2'' \to 0$ to determine that k_{2z}^R in the lossless case is along the positive imaginary axis. (b) For a slightly "gainy vacuum" ($\epsilon_2'' < 0$) the two solutions for k_{2z} are in the second and fourth quadrants, and k_{2z}^R approaches the negative imaginary axis as $\epsilon_2'' \to 0$. The magnitudes of the real and imaginary parts of k_{2z} in (a) and (b) are approximated using the first order Taylor expansion for small ϵ_2'': $k_0^2|\epsilon_2''| \ll k_x^2 - k_0^2$.

To have an unambiguous labeling convention for the case $\epsilon_2'' = 0$, we emphasize that one must take the limit as *loss* approaches zero, which can be different from the limit as gain approaches zero in the case of evanescent waves.

Finally, it is worth noting that this discontinuity in the two limiting cases, apart from being a footnote in establishing a labeling convention, is actually at the heart of the debate over single-surface amplified TIR. When medium two has gain, if one chooses k_{2z}^R as the transmitted wavevector (in accordance with our postulate), then it seems unphysical that as the gain approaches (but does not reach) zero the transmitted wave should still be strongly amplified. To remedy this situation it has been suggested that the correct choice for the transmitted wavevector should be k_{2z}^L when medium two has gain and $k_x > k_0$, so that the transmitted wave decays in the $+z$-direction. We believe, instead, that the discontinuity in the two limits is not as unphysical as it might appear at first: the transmitted wave propagates a large distance in the x-direction while barely moving forward in the z-direction (since $k_x \gg \mathrm{Re}(k_{2z}^R)$), so the large gain in the z-direction is actually a result of the long propagation distance along the x-direction. Far more unphysical, in our opinion, is the decision to switch the transmitted wavevector from k_{2z}^R when $k_x < k_0$ to k_{2z}^L when $k_x > k_0$.

All of these arguments aside, however, the purpose of our paper has been to demonstrate the Fresnel mechanism by which the specularly reflected beam from a finite-thickness slab is amplified, both below and above the critical angle.

II. GAIN SATURATION

Physicists familiar with the principles of lasers should be rightfully wary of a steady-state solution with roundtrip coefficient $|\nu|$ greater than one. When an active medium is pumped strongly enough to generate a sufficiently large population inversion to yield $|\nu|$ greater than one, light initially generated by spontaneous emission in the cavity will be amplified after each roundtrip. However, the field amplitude does not grow without bound–as the field gains strength the upper state lifetime is reduced by stimulated emission, which causes the population inversion to decrease to a level such that $\nu = 1$, resulting in steady-state lasing. This gain reduction with increasing field amplitude is known as gain saturation. In a laser, therefore, the situation $|\nu| > 1$ is only a transient state. It clearly cannot be a steady-state solution, because the field would grow without bound.

The situation changes when we allow an incident wave to strike the active medium, as we do in this paper. Note that ν is defined as the roundtrip coefficient *in the absence of an incident wave*; that is, the reflectivity r_{21} is calculated by assuming that there is no wave in medium one arriving at the cavity. To account for the incident wave, we can define an effective facet reflectivity at the two-one interface $r_{21}^{\text{eff}} \equiv E_2^R/E_2^L$. Furthermore, we can define an effective roundtrip coefficient in the slab which replaces r_{21} with r_{21}^{eff}, that is, $\nu^{\text{eff}} = r_{21}^{\text{eff}} r_{23} \exp(2ik_{2z}^R d)$. We emphasize that *every possible steady-state solution to the problem under consideration, whether the slab is passive or active, and whether there is an incident wave or not, satisfies the condition* $\nu^{\text{eff}} = 1$. This is a fundamental property of steady-state solutions: the field in the slab must regenerate itself after every roundtrip, once all sources and sinks have been accounted for. Therefore, in situations where $|\nu| > 1$, the incident wave must, upon transmission into medium two, interfere destructively with the circulating field in the slab so that $|r_{21}^{\text{eff}}| < |r_{21}|$, which ultimately forces ν^{eff} toward 1. In summary, when there is no incident wave the situation $|\nu| > 1$ is temporary because the field will grow until gain saturation (a nonlinear effect) forces the $\nu = 1$ solution. With an incident wave, a linear steady-state solution is possible even when $|\nu| > 1$ because of the

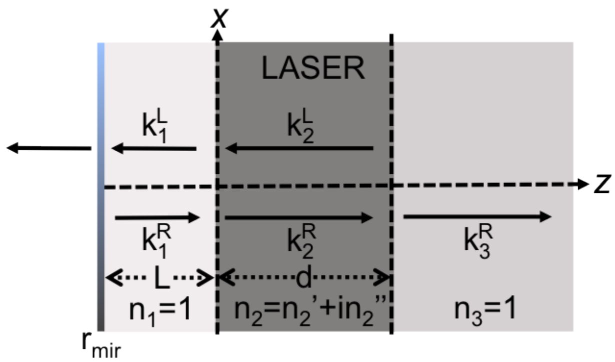

FIG. S2. Schematic of a laser with self-feedback. A mirror of reflectivity r_{mir} allows light emitted by medium two to be re-injected. The feedback alters the laser threshold, and can be modeled by an effective facet reflectivity r_{21}^{eff}, which differs from the bare facet reflectivity r_{21} in the absence of feedback.

reduction in the effective facet reflectivity r_{21}^{eff}, which prevents the unbounded growth of the fields so that one does not have to rely on gain saturation to avoid a nonphysical divergence.

An example that illustrates well the distinction between ν and ν^{eff} is a laser subject to self-feedback, shown in Fig. S2. In this case, the laser (medium 2) emits light in a direction normal to both of its facets, and a mirror of reflectivity r_{mir} placed a distance L from the left facet reflects some of the laser output and creates the wavevector $\mathbf{k}_1^R$, which is then reinjected into the laser. (To be clear, this is different from the situation considered in our paper, for which $\mathbf{k}_1^R$ is generated by an external source.) For a suitable choice of L and r_{mir}, the re-injected wave will interfere destructively with the circulating field in the gain medium, effectively reducing the reflection coefficient at the 2-1 interface without altering its phase. Specifically, for $n_1 = 1$ and $n_2 = n_2' + in_2''$, the facet reflectivity in the absence of feedback is $r_{21} = (n_2 - 1)/(n_2 + 1)$. The feedback from the external mirror will reduce the effective reflection coefficient to $r_{21}^{\text{eff}} = \alpha r_{21}$, where α is real and satisfies $0 \le \alpha \le 1$, provided r_m and L are chosen such that

$$r_m \exp(i4\pi L/\lambda_o) = \frac{(1-\alpha)r_{21}}{\alpha r_{21}^2 - 1}. \tag{4}$$

So long as $|r_{21}| < 1$ (which is generally the case), it is possible to choose r_m and L so that $|r_m| < 1$, i.e., the external mirror is a simple, passive component. As an application of Eq. 4, consider the case $n_2 = 1.5 - 0.01i$ and $\lambda_o = 1$ μm. The bare facet reflectivity is $r_{21} = 0.2\exp(-0.016i)$. If we want the effective facet reflectivity to be $r_{21}^{\text{eff}} = 0.1\exp(-0.016i)$

(half the magnitude but the same phase as the bare facet reflectivity), we choose $\alpha = 0.5$ and find that we should place a mirror of reflectivity $r_m = -0.102$ a distance $L = 0.499$ μm from the left facet.

Because the external mirror reduces the effective facet reflectivity so that $|r_{21}^{\text{eff}}| < |r_{21}|$, the effective roundtrip coefficient $|\nu^{\text{eff}}|$ will be less than $|\nu|$. When medium two is pumped beyond the lasing threshold, ν^{eff} will be clamped to 1, which means that $|\nu|$ will be greater than 1. There also exists a subthreshold pumping regime for which $|\nu| > 1$. The example of the laser with self-feedback demonstrates that an active cavity subject to an incident field can have a roundtrip coefficient ν with magnitude greater than 1, and that this situation is neither transient nor unstable.

In the above example, there is always a well-defined phase relationship between the field circulating in the laser and the re-injected field. This will not necessarily be true when the incident wave is generated by an external source: spontaneous emission, being a stochastic process, can give rise to a field within medium two that has no well-defined phase relationship with the incident wave. Whether this leads to instabilities remains to be decisively answered, but we note that the existing experimental evidence in the $|\nu| > 1$ regime has not, to our knowledge, detected any instabilities [14, 20, 21]. It is also worth commenting on an important difference between spontaneously-emitted cavity-photons traveling parallel to the surface-normal of the cavity facets versus those emitted at an oblique angle to the optical axis. For photons emitted parallel to the surface-normal, the slab is clearly a cavity that provides feedback, as the light retraces its path on every roundtrip. However, as mentioned in the main text, the roundtrip coefficient is a function of incidence angle, and for the material parameters used for Fig. 2(b), $|\nu|$ exceeds one only for incidence angles $\theta > 27.43°$. In particular, this means that $|\nu| < 1$ for $\theta = 0$; therefore, the gain is insufficient for a spontaneously emitted cavity photon in the $\theta = 0$ direction to cause lasing. In contrast, a cavity photon emitted spontaneously at a large θ such that $|\nu| > 1$ would seem to experience net amplification after each roundtrip. (We say 'seem to' because this is the intuitive interpretation to us; however, we mention again that the experiments that have probed the regime of $|\nu| > 1$ have not detected such problems with spontaneous emission [14, 20, 21].) Should these obliquely-traveling spontaneously-emitted photons prove problematic in a future experiment, however, they will anyway exit the slab at the top and bottom facets, since any real slab must have a finite length in the x-direction. One could also coat the top and bottom facets with broadband antireflection coatings to facilitate this removal; this way, these spontaneously emitted photons would leave the slab before they are amplified to the point where they saturate the gain. We mention this only as a consideration for a potential experiment that looks for the pre-excitation mechanism. Our intent in this Letter has been to explore the predictions of the Fresnel solution for a slab with $|\nu| > 1$; in the end, only experiments can decide whether this solution is physical.

III. PULSE OF LIGHT INCIDENT ON GAINY SLAB WITH $|\nu| > 1$

The video file pulse_video.avi included online is a time-lapse video of $|E_y|^2$ of a pulse of light, rather than a beam, for the same material parameters as in Fig. 2(b) of the main text: $\epsilon_1 = \epsilon_3 = 2.25$, $\epsilon_2 = 1 - 0.01i$, $\mu_1 = \mu_2 = \mu_3 = 1$, $d = 28$ μm. The white vertical lines in the video identify the 1-2 and 2-3 interfaces. The incident pulse is s-polarized and Gaussian in both space (FWHM = 13.3 μm) and time (FWHM = 50 fs, or 15 optical cycles). The central wavelength of the pulse is $\lambda_o = 1$ μm, and the mean incidence angle (i.e., averaged over all constituent planewaves) is 30°. The size of each video frame is 210 μm by 150 μm (height by width). The time elapsed between frames is 10 fs, and the entire video spans 1.22 ps (123 frames total). The field $|E_y|^2$ is plotted on a logarithmic scale covering 3 decades, i.e. red corresponds to the maximum intensity and blue corresponds to intensities less than or equal to 1/1000th of the maximum. The background in this image is blue, which corresponds to the minimum of $|E_y|^2$, whereas the background in Figs. 2(a) and 2(b) is green because it is the field E_y that was plotted in that case, so that blue corresponded to the maximum negative field.

In the video, one first sees the incident pulse near the bottom left of the screen, traveling up and to the right. The pre-excitation is soon seen in the slab at the bottom of the frame, and the reflected pulse that corresponds to the $m = 1$ term in the primed partial wave expansion leaves the slab and propagates up and to the left in medium one. The pre-excitation in the slab then undergoes one roundtrip as it zig-zags upward, giving rise to a transmitted pulse in medium three followed by the $m = 0$ reflected pulse in medium one. The pre-excitation then makes one more roundtrip, giving rise to another transmitted pulse in medium three, and then approaches the two-one interface at the same time the incident pulse arrives from the opposite side. The two pulses interfere in such a way as to yield an amplified specularly reflected pulse by entirely depleting the energy content of the slab. The fact that the pre-excitation in the slab travels in the $+x$-direction clearly distinguishes this behavior from negative refraction.

IV. DESCRIPTION OF SIMULATION

The E-field plots of the Gaussian beams and the video of the pulse were created using MATLAB. The field at each pixel is determined by superposing a large (but of course finite) number of planewave solutions. Therefore, the plots represent analytical solutions to Maxwell's equations.

As described in the main text, the response of the slab to an incident s-polarized planewave with amplitude E_1^R and wavevector $\boldsymbol{k}_1^R = k_x\hat{\boldsymbol{x}} + k_{1z}^R\hat{\boldsymbol{z}}$ is given by

$$E_y(x,z) = \begin{cases} E_1^R \exp(ik_x x + ik_{1z}^R z) + E_1^L \exp(ik_x x + ik_{1z}^L z) & : z \le 0 \\ E_2^R \exp(ik_x x + ik_{2z}^R z) + E_2^L \exp(ik_x x + ik_{2z}^L z) & : 0 \le z \le d \\ E_3^R \exp[ik_x x + ik_{3z}^R (z-d)] & : z \ge d \end{cases} \tag{5}$$

and the time-dependence factor $\exp(-i\omega t)$ is not explicitly written. The wavevector components k_{2z}^R and k_{3z}^R are determined by the dispersion relation

$$k_{\ell z}^R = \sqrt{(\omega/c)^2 \mu_\ell \epsilon_\ell - k_x^2}, \tag{6}$$

where μ_ℓ and ϵ_ℓ are the relative magnetic permeability and electric permittivity constants of material ℓ, and the sign of the square root is chosen according to the prescription described in Supplementary Sec. 1. The four unknown wave amplitudes are found by satisfying Maxwell's boundary conditions to be

$$E_2^R = \frac{2k_{1z}^R(k_{3z}^R + k_{2z}^R)E_1^R}{(k_{2z}^R + k_{1z}^R)(k_{3z}^R + k_{2z}^R) + \exp(2ik_{2z}^R d)(k_{3z}^R - k_{2z}^R)(k_{2z}^R - k_{1z}^R)} \tag{7}$$

$$E_2^L = \frac{-2k_{1z}^R(k_{3z}^R - k_{2z}^R)E_1^R}{(k_{2z}^R - k_{1z}^R)(k_{3z}^R - k_{2z}^R) + \exp(-2ik_{2z}^R d)(k_{3z}^R + k_{2z}^R)(k_{2z}^R + k_{1z}^R)} \tag{8}$$

$$E_1^L = E_2^R + E_2^L - E_1^R \tag{9}$$

$$E_3^R = E_2^R \exp(ik_{2z}^R d) + E_2^L \exp(-ik_{2z}^R d). \tag{10}$$

To construct the Gaussian beam from the planewave solutions, we begin by expressing E_y in the $z = 0$ plane for a beam traveling parallel to the z-axis

$$E_y(x, z = 0) = E_0 \exp\left(-\frac{x^2}{2\sigma_x^2}\right), \tag{11}$$

where E_0 is the peak amplitude and σ_x is directly proportional to the spatial FWHM

$$w_x = 2\sqrt{2\ln 2}\sigma_x. \tag{12}$$

By Fourier transforming and subsequently inverting the transform, the field can equivalently be written as an integral in k-space,

$$E_y(x, z = 0) = \int_{-\infty}^{\infty} dk_x E_1^R(k_x) \exp(ik_x x), \tag{13}$$

where

$$E_1^R(k_x) = \frac{E_0\sigma_x}{\sqrt{2\pi}} \exp\left(\frac{-k_x^2}{2(1/\sigma_x)^2}\right), \tag{14}$$

and the FWHM in k-space is

$$w_k = 2\sqrt{2\ln 2}/\sigma_x. \tag{15}$$

To propagate the beam beyond the $z = 0$ plane, we associate with each value of k_x a component k_{1z}^R such that the total wavevector obeys the dispersion relation in medium one,

$$k_{1z}^R(k_x) = \sqrt{(\omega/c)^2\mu_1\epsilon_1 - k_x^2}. \tag{16}$$

Now the Gaussian beam can be expressed as a function of x and z by

$$E_y(x, z) = \int_{-\infty}^{\infty} dk_x E_1^R(k_x) \exp[i(k_x x + k_{1z}^R z)]. \tag{17}$$

At this point, we must approximate the integral in Eq. 17 by discretization so that the calculation can be carried out by a computer. We restrict k_x to a finite sampling width w_s given by $-w_s/2 \leq k_x \leq w_s/2$, and sample the beam equidistantly within this region with a total number of samples N_s. The integral in Eq. 17 is approximated by the sum

$$E_y(x, z) = \sum_{k_x=-w_s/2}^{w_s/2} \Delta k_x E_1^R(k_x) \exp[i(k_x x + k_{1z}^R z)], \tag{18}$$

where

$$\Delta k_x = \frac{w_s}{N_s - 1}. \tag{19}$$

At this point, it is helpful to think of E_1^R, k_x, and k_{1z}^R as vectors containing N_s numerical elements each. To rotate the beam so that it travels at an angle θ to the z-axis, we perform the transformation

$$k_x \rightarrow \cos(\theta)k_x + \sin(\theta)k_z \tag{20}$$

$$k_{1z}^R \rightarrow -\sin(\theta)k_x + \cos(\theta)k_{1z}^R \tag{21}$$

on each element of k_x and k^R_{1z}. (The Fourier amplitude of each plane-wave $E^R_1(k_x)$ is unaffected by the rotation in the case of s-polarized light.) Finally, to displace the waist of the beam to some location (x_0, z_0) in the incidence medium, one must multiply each Fourier amplitude by

$$E^R_1(k_x) \to E^R_1(k_x)\exp[-i(k_x x_0 + k^R_{1z} z_0)]. \tag{22}$$

With these redefined values for E^R_1, k_x, and k^R_{1z}, the sum in Eq. 18 is a good approximation to a Gaussian beam traveling at an angle θ whose waist is located at (x_0, y_0). The total E-field at any point in the system is given by

$$E_{\mathrm{tot}}(x,z) = \begin{cases} \mathrm{Real}\{\sum \Delta k_x \left(E^R_1(k_x)\exp[i(k_x x + k^R_{1z} z)] + E^L_1(k_x)\exp[i(k_x x + k^L_{1z} z)]\right)\}, & z \le 0 \\ \mathrm{Real}\{\sum \Delta k_x \left(E^R_2(k_x)\exp[i(k_x x + k^R_{2z} z)] + E^L_2(k_x)\exp[i(k_x x + k^L_{2z} z)]\right)\}, & 0 \le z \le d \\ \mathrm{Real}\{\sum \Delta k_x E^R_3(k_x)\exp[i(k_x x + k^R_{3z} z)]\}, & z \ge d \end{cases} \tag{23}$$

where E^L_1, E^R_2, E^L_2, and E^R_3 are calculated element-wise from $E^R_1(k_x)$ according to Eqs. 7-10. The beam plots in Fig. 2 of the main text are calculated pixel-by-pixel from the sum in Eq. 23, with the values of x and z indicating the location of the pixel. The resultant field is normalized to the maximum field value in the image, and displayed in color. The pulse video is calculated similarly, except that the field is Gaussian in space and time, and so the field must be sampled in both the spatial and temporal frequency domains. The calculation time is significantly longer for the pulse compared to the beam, and the simulations are only practical to run on a supercomputer.

The finite nature of the sampling has consequences which must be considered in order to be sure that our results are not affected by numerical artifacts. Firstly, the truncation of the Gaussian beam in k-space to the sampling width w_s leads to a convolution with a sinc function in the spatial domain. Therefore, the side-tail of our beam is not truly Gaussian; rather, the envelope of the side-tail is Gaussian but the side-tail itself exhibits periodic sinc-like fluctuations in intensity (which cannot be seen in Fig. 2 of the main text, but can be seen in logarithmic plots which resolve the small intensities of the side-tail). The sampling width chosen for Fig. 2 was $w_s = 2w_k$ (with $N_s = 501$). We made sure that other choices of the sampling width, $w_s = 3w_k$ and $4w_k$ (with proportionally larger N_s so that Δk_x remained constant), did not affect the behavior of the plots. Therefore, our conclusions are not affected by the precise value of the sampling width w_s. Secondly, the finite number of samples N_s

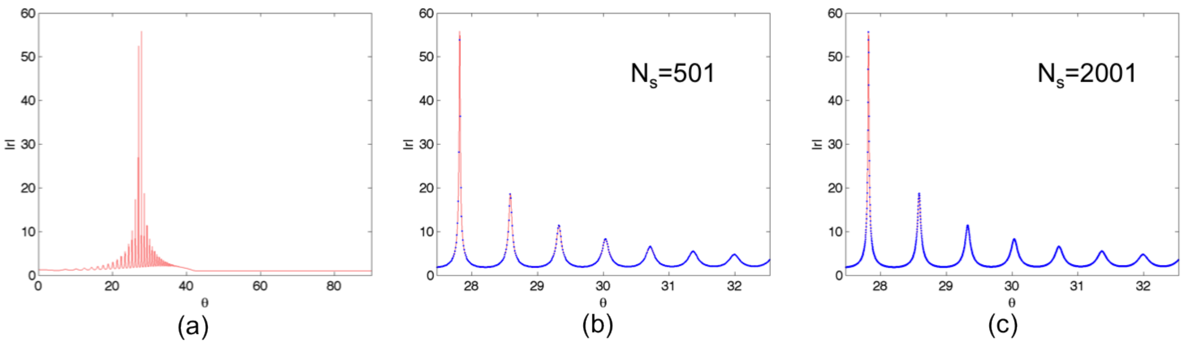

FIG. S3. A pictorial depiction of the Fourier domain sampling used to generate Fig. 2(b) of the main text. (a) A plot of $|r|$ vs. incidence angle θ for the case of $\lambda_o = 1$ μm, ϵ_1=ϵ_3=2.25, $\epsilon_2 = 1 - 0.01i$, and $d = 28$ μm. For a beam-width FWHM of 13.3 μm, the sampling width $w_s = 2w_k$ that we chose restricted the range of incidence angles of the planewaves used to construct the beam to $27.47° < \theta < 32.53°$. In plots (b) and (c), the blue dots overlaying the reflectivity plot indicate the incidence angles of the planewaves that were summed over for (b) $N_s = 501$ samples and (c) $N_s = 2001$ samples. The plot in Fig. 2(b) of the main text looks visually identical for both choices of sampling values.

implies the spectrum of k_x values is discrete, so the incident beam is periodic in space. This means that in the plots of Fig. 2 in the main text, there is not just one incident beam but an infinite number of them impinging on the slab, spaced periodically along the x-axis by a distance $2\pi/\Delta k_x = 2830$ μm. If the sampling is increased from $N_s = 501$ to 2001 while keeping $w_s = 2w_k$ constant (see Fig. S3), the distance between adjacent beams increases to 11330 μm, but the plots in both Figs. 2(a) and 2(b) of the main text look identical to the ones with 501 samples. Therefore, 501 samples is sufficient in this case to ensure that the (periodically repeated) beams do not interfere with each other, and that the plot is a good representation of the field of a single beam.

[1] J.B. Pendry. "Negative refraction makes a perfect lens." Phys. Rev. Lett. **85**, 3966-69 (2000).

[2] D.R. Smith, D. Schuring, M. Rosenbluth, S. Schultz, S.A. Ramakrishna, and J.B. Pendry. "Limitations on subdiffraction imaging with a negative refractive index slab." Appl. Phys. Lett. **82**, 1506-1508 (2003).

[3] N. Garcia and M. Nieto-Vesperinas. “Left-Handed Materials Do Not Make a Perfect Lens.” Phys. Rev Lett. **88**, 20, 207403 (2002).

[4] X. Jiang, Q. Li and C.M. Soukoulis. “Symmetry between absorption and amplification in disordered media.” Phys. Rev. B **19**, 9007 (1999).

[5] S.A. Ramakrishna. “On the dual symmetry between absorbing and amplifying random media.” Pramana **62**, 1273-1279, (2004).

[6] J.C.J. Paasschens, T. Sh. Misirpashaev and C.W.J. Beenakker. “Localization of light: dual symmetry between absorption and amplification.” Phys. Rev. B **54**, 11887 (1996).

[7] P. Pradhan and N. Kumar. “Localization of light in coherently amplifying random media.” Phys. Rev. B **50**, 9644 (1994).

[8] Z.Q. Zhang. “Light amplification and localization in randomly layered media with gain.” Phys. Rev. B **52**, 7960 (1995).

[9] Y. Chen, P. Fischer, and F.W. Wise. “Negative Refraction at Optical Frequencies in Nonmagnetic Two-Component Molecular Media.” Phys. Rev. Lett. **95**, (2005).

[10] S.A. Ramakrishna. “Comment on ‘Negative Refraction at Optical Frequencies in Nonmagnetic Two-Component Molecular Media.’” Phys. Rev. Lett. **98**, (2007).

[11] S.A. Ramakrishna and O.J.F. Martin. “Resolving the wave vector in negative index media.” Opt. Lett. **30**, 19, 2626-2628 (2005).

[12] J.B. Geddes III, T.G. Mackay, and A. Lakhtakia. “On the refractive index for a nonmagnetic two-component medium: Resolution of a controversy.” Opt. Comm. **280**, 120-125 (2007).

[13] B. Nistad and J. Skaar. “Causality and elecromagnetic properties of active media.” Phys. Rev. E **78**, 036603 (2008).

[14] C.J. Koester. “Laser action by enhanced total internal reflection.” IEEE J. Quantum Electron. **QE-2**, 580-4 (1966)

[15] P.R. Callary and C.K. Carniglia. “Internal reflection from an amplifying layer.” J. Opt. Sci. Am. **66**, 8 (1976).

[16] J. Fan, A. Dogariu, and L.J. Wang. “Amplified total internal reflection.” Opt. Express **11**, 299-308 (2003).

[17] K.J. Willis, J.B. Schneider, and S.C. Hagness. “Amplified total internal reflection: theory, analysis, and demonstration of existence via FDTD.” Opt. Express **16**, 1903-14 (2008).

[18] A. Siegman. “Fresnel Reflection and Evanescent Gain.” OPN, 38–45 (Jan. 2010).

[19] J.O. Grepstad and J. Skaar. "Total internal reflection and evanescent gain." Opt. Express **19**, 21404 (2011).

[20] E.P. Ippen and C.V. Shank. "Evanescent-field-pumped dye laser." Appl. Phys. Lett. **21**, 301 (1972).

[21] H.H. Choi, H.J. Kim, J. Noh and W. Jhe. "Amplified total internal reflection from a plane dielectric boundary." Jour. Korean Phys. Soc. **42**, 80 (2003).

Fourth Report of Referee A, received December 19th, 2013:

I have re-read the manuscript and authors' response to my previous criticism. Unfortunately, I cannot support the publication of this manuscript in Phys. Rev. Lett.

As before, I consider the very approach of pre-excitation presented by the authors to be essentially wrong because it takes Fresnel formalism beyond the limits of its applicability.

One of the few exceptions to this rule is the case of amplified total internal reflection, primarily because in this particular case one can avoid round-trip gain. I have tried to repeat the calculations that authors presented in their reply. Although in my calculations I was not able to reproduce the curves exactly, I assume that the relatively minor deviations between authors' and my calculations come from a possible typo in the number of 9-s either in authors' reply or in my setup. I found that the reflectivity curves in Fig.S1b strongly depend on polarization and, more importantly, on the choice of the kz sign. Note that the authors are correct that the sign of kz is not important for second layer in the system; however, it becomes of primary importance for the third layer (that has gain).

In my calculations it is seen that if we cut the complex plane along negative real axis (I presume this is the cut that authors use), the real-space field distribution indicates pre-excitation proposed by the authors. At the same time, if one cuts complex plane along negative imaginary axis, the whole pre-excitation dynamics disappears. I want to point out that in this latter case there is no round-trip gain in the system. Both waves in layer 2 and a single wave in layer 3 assumed to be decaying from their interfaces.

I therefore remain convinced that it is unphysical to use Fresnel formalism in a system with round-trip gain and that one should not blindly trust the results obtained from solutions of mathematical equations in this limit. My impression is that this type of manuscript would have been appropriate for a specialized journal which publishes articles with preliminary data. However, I would not recommend a publication of this work in The Physical Review. Needless to say, I would not recommend publication of this work in Phys. Rev. Lett.

Final decision of editor, sent December 19^{th}, 2013:

Dear Dr. Mansuripur,

The above manuscript has been reviewed by our referee.

A critique drawn from the report appears below. On this basis, we judge that the paper is not appropriate for Physical Review Letters, but might be suitable for publication in another journal, possibly with some revision. Therefore, we recommend that you submit your manuscript elsewhere. In accordance with our standard practice (see memo appended further below), this concludes our review of your manuscript.

Yours sincerely,

Saad E. Hebboul
Senior Assistant Editor
Physical Review Letters

Dr. Samuel D. Bader
Associate Editor
Applied Physics Letters

March 6, 2014

Dear Dr. Bader,

Thank you for agreeing to evaluate our manuscript for publication in *Applied Physics Letters*. In the manuscript we are submitting to *APL*, we have made a few changes relative to the final submission that was rejected at *Physical Review Letters*. These changes are highlighted in the accompanying pdf file. The theory of a cavity with net roundtrip gain also applies to evanescent waves in the negative-index lens ("Pendry's lens"). While we had mentioned this connection briefly in the PRL version, here we expand the argument in order to demonstrate that this theory is applicable to passive as well as active media. In addition, we have included references to literature on the negative-index lens that was published in APL.

Sincerely,

____

Tobias S. Mansuripur
Graduate Student
Physics Department
Harvard University

Manuscript submitted to Applied Physics Letters (March 6, 2014).
Changes relative to version 4 of PRL submission are highlighted.

Fresnel reflection from a cavity with net roundtrip gain

Tobias S. Mansuripur[1,*] and Masud Mansuripur[2]
[1]Department of Physics, Harvard University, Cambridge, MA 02138
[2]College of Optical Sciences, The University of Arizona, Tucson, AZ 85721

A planewave incident on an active etalon with net roundtrip gain may be expected to diverge in field amplitude, yet applying the Fresnel formalism to Maxwell's equations admits a convergent solution. We describe this solution mathematically, and provide additional insight by demonstrating the response of such a cavity to an incident beam of light. Cavities with net roundtrip gain have often been overlooked in the literature, and a clear understanding of their behavior yields insight to negative refraction in nonmagnetic media, a duality between loss and gain, amplified total internal reflection, and the negative-index lens.

The Fresnel coefficients govern the reflection and transmission of light for the simplest possible scenarios: at planar interfaces between homogeneous media. Despite this simplicity, some interesting solutions have been discovered only recently, such as 1) the amplification of evanescent waves in a passive, negative-index slab [1–5] and 2) a duality between loss and gain leading to the localization of light in both cases [6–10]. In addition, controversy regarding the proper choice of the wavevector in active media has persisted in relation to the possibility of 3) negative refraction in nonmagnetic media [11–15] as well as 4) single-surface amplified total internal reflection (TIR) [16–21]. It turns out that all four of these cases share a common thread: the presence of a cavity whose roundtrip gain exceeds the loss. In this Letter, we explore in detail the Fresnel solution for such a cavity. We find that its peculiar properties help us to understand the four aforementioned phenomena, and could also enable novel device functionalities.

To begin, we establish a convention that allows us to more clearly discuss the direction of energy flow. For the single-surface problem, shown in Fig. 1(a), the incident wavevector in medium one is $\boldsymbol{k}_1^R = k_x\hat{\boldsymbol{x}} + k_{1z}^R\hat{\boldsymbol{z}}$, and the reflected wavevector is $\boldsymbol{k}_1^L = k_x\hat{\boldsymbol{x}} + k_{1z}^L\hat{\boldsymbol{z}}$, where $k_{1z}^L = -k_{1z}^R$. The superscript R (L) indicates that the wave carries energy to the right (left)–in other words, that the time-averaged z-component of the Poynting vector is positive (negative). The real-valued component k_x, once determined by the incident wave, is the same for all wavevectors in the system. For the transmitted wavevector, the dispersion relation offers two choices for k_{2z},

$$k_{2z} = \pm\sqrt{(\omega/c)^2\mu_2\epsilon_2 - k_x^2}, \qquad (1)$$

where ω is the angular frequency, c is the speed of light in vacuum, and μ_2 and ϵ_2 are the relative permeability and permittivity. It is universally agreed that the correct choice for k_{2z} in the single-surface problem is k_{2z}^R (i.e., that the transmitted energy flows away from the interface), irrespective of the material parameters or the nature of the incident wave, except possibly in the case of amplified TIR, for which there remains debate. Due to this controversy, let us postulate for now that k_{2z}^R is the correct choice in all cases, so that we can unambiguously define the single-surface Fresnel reflection and transmission coefficients

$$r_{\ell m} = \frac{\tilde{k}_{\ell z}^R - \tilde{k}_{mz}^R}{\tilde{k}_{\ell z}^R + \tilde{k}_{mz}^R}, \; t_{\ell m} = \frac{2\tilde{k}_{\ell z}^R}{\tilde{k}_{\ell z}^R + \tilde{k}_{mz}^R} \qquad (2)$$

where we have generalized the result for incidence medium ℓ and transmission medium m. For s-polarization we have defined $\tilde{k}_{nz} \equiv k_{nz}/\mu_n$, while for p-polarization $\tilde{k}_{nz} \equiv k_{nz}/\epsilon_n$. (In cases where both choices for k_{2z} result in no energy flow in the z-direction, such as for evanescent waves in a transparent medium, our prescription is to add a small amount of loss to the slab which will unambiguously distinguish k_{2z}^R and k_{2z}^L, then take the limit as the loss goes to zero [22].)

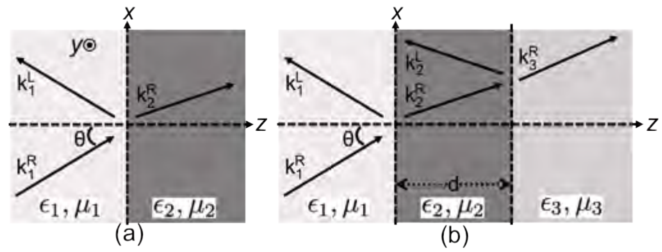

FIG. 1. Geometry of the (a) single-surface and (b) cavity problems. All media are infinite in the x and y-directions. The arrows denote the wavevectors present in each layer.

We now consider the case of light incident on a cavity, shown in Fig. 1(b). The total E-field resulting from an s-polarized incident wave in medium one with amplitude E_1^R is given by

$$E_y(x,z) = \begin{cases} E_1^R \exp(ik_x x + ik_{1z}^R z) \\ \quad +E_1^L \exp(ik_x x + ik_{1z}^L z) & : z \le 0 \\ E_2^R \exp(ik_x x + ik_{2z}^R z) \\ \quad +E_2^L \exp(ik_x x + ik_{2z}^L z) & : 0 \le z \le d \\ E_3^R \exp[ik_x x + ik_{3z}^R (z-d)] & : z \ge d \end{cases} \qquad (3)$$

where the time-dependence factor $\exp(-i\omega t)$ has been omitted. The most direct route to solve for the four un-

* mansuripur@physics.harvard.edu

known wave amplitudes is to enforce Maxwell's boundary conditions at $z = 0$ and $z = d$, which yields four equations that can be solved for the four unknowns. The resulting reflection coefficient from the slab can be expressed in terms of the single-surface Fresnel coefficients as

$$r \equiv \frac{E_1^L}{E_1^R} = \frac{r_{12} + r_{23}\exp(2ik_{2z}^R d)}{1-\nu} \tag{4}$$

where

$$\nu = r_{21}r_{23}\exp(2ik_{2z}^R d) \tag{5}$$

is referred to as the roundtrip coefficient; the amplitude of a planewave circulating in the slab is multiplied by this factor after each roundtrip in the absence of any sources outside the slab. (Although we explicitly discuss s-polarized light, our conclusions as well as Eqs. 4 and 5 hold for both polarization states.) We emphasize that the reflection coefficient given by Eq. 4 is a valid solution to Maxwell's equations for any value of ν. The roundtrip coefficient ν has an important physical meaning, and intuitively one would expect three different regimes of behavior when the magnitude of ν is less than, equal to, or greater than one. The case where $|\nu| < 1$ governs passive slabs (in most but not all cases) and sufficiently weakly amplifying slabs. When $\nu = 1$ the slab behaves as a laser and emits light even in the absence of an incident wave, which manifests itself mathematically as an infinitely large reflection amplitude. The case where $|\nu| > 1$, however, is rarely openly acknowledged [6, 15, 17].

Probably the reason for the neglect of the $|\nu| > 1$ steady-state solution is the seemingly intuitive assumption that the fields should diverge when there is net roundtrip gain. (See the supplemental material for a discussion of gain saturation and lasing [22].) This assumption is only reinforced by examining a second well-known solution method for the reflection coefficient that decomposes the reflected wave amplitude E_1^L into a sum over partial waves, yielding the reflection coefficient

$$r = r_{12} + t_{12}t_{21}r_{23}\exp(2ik_{2z}^R d)\sum_{m=0}^{\infty}\nu^m. \tag{6}$$

Heuristically, the first term r_{12} (hereinafter referred to as the "specular" partial wave) of Eq. 6 results from the single-surface reflection of the incident wave at the 1-2 interface, and the geometric series accounts for the contributions to the reflected wave following multiple roundtrips within the slab. When $|\nu| < 1$, the geometric series in Eq. 6 converges to $(1-\nu)^{-1}$, giving the same result as found by matching the boundary conditions in Eq. 4. When $|\nu| > 1$, however, the geometric series diverges and the reflection coefficient is infinite. Intuitively, this divergence seems reasonable, since we expect any light that couples into a slab with $|\nu| > 1$ to be amplified after each roundtrip, and therefore grow without bound. Nevertheless, Eq. 4 yields a finite reflection coefficient even when $|\nu| > 1$, so how can we reconcile these two very different solutions?

In fact, the partial wave method *can* be used to find the $|\nu| > 1$ convergent solution. First, one can check that the reflection coefficient given by Eq. 4 is invariant under the transformation $k_{2z}^R \leftrightarrows k_{2z}^L$. (This can be interpreted simply as a relabeling of the waves $E_2^R \leftrightarrows E_2^L$ in Eq. 3 that does not affect the final result.) Applying this same transformation to the partial wave sum of Eq. 6 [17], we can express the reflection coefficient as

$$r = r'_{12} + t'_{12}t'_{21}r'_{23}\exp(2ik_{2z}^L d)\sum_{m=0}^{\infty}\nu'^m, \tag{7}$$

where the prime indicates the substitution $k_{2z}^R \to k_{2z}^L$. Because the new roundtrip coefficient, $\nu' = r'_{21}r'_{23}\exp(2ik_{2z}^L d)$, is equal to ν^{-1}, in cases where $|\nu| > 1$ the primed partial wave sum of Eq. 7 will converge to the reflection coefficient of Eq. 4. (This duality between ν and ν^{-1} provides a simple mathematical explanation for the loss/gain duality observed by others [6–10]).

The physical implications of the substitution $k_{2z}^R \to k_{2z}^L$ in the partial wave sum can best be seen by examining the behavior of a "finite-diameter" beam of light incident obliquely on the slab. By numerically superposing a finite number of planewave solutions to Eq. 3 [22] with appropriate amplitudes and incidence angles in the range $27.47° < \theta < 32.53°$, we create a Gaussian (to within the sampling accuracy) beam incident on the slab at 30° with a full-width at half-maximum (FWHM) beam-diameter of 13.3 μm. All media are nonmagnetic, and we choose $\epsilon_1 = \epsilon_3 = 2.25$ and the slab to be an amplifying medium with $\epsilon_2 = 1 - 0.01i$. The free-space wavelength of the beam is $\lambda_o = 1$ μm. We can examine the transition from $|\nu| < 1$ to $|\nu| > 1$ simply by varying d, since both $|r_{21}|$ and $|r_{23}|$ are less than one (and independent of d), whereas $|\exp(2ik_{2z}^R d)|$ (and hence ν) increases monotonically with d (because k_{2z}^R has a negative imaginary part). A plot of the field $E_y(x, z)$ at one instant of time is shown in Fig. 2(a) for $d = 19$ μm, which was chosen so that $|\nu|$ is slightly less than one for all constituent planewaves of the beam ($0.46 < |\nu| < 0.99$). The arrows overlying the plot point in the direction of the time-averaged Poynting vector within their vicinity, indicating the direction of energy flow in the system, and the incident beam is uniquely identified by the white arrow. The beam behaves as we expect it to: the incident beam strikes the slab near ($x = 0$, $z = 0$), giving rise to a specularly reflected beam as well as a refracted beam that 'zig-zags' up the slab, which in turn generates a reflected beam in medium one each time it strikes the 2-1 interface. (The field amplitude is plotted on a linear scale, and so the incident beam as well as the specularly reflected beam appear faint relative to the subsequently amplified portions of the beam.) Each of these reflected beams can intuitively be associated with a term of the partial wave expansion of Eq. 6–either the specular term or the mth term of the geometric series.

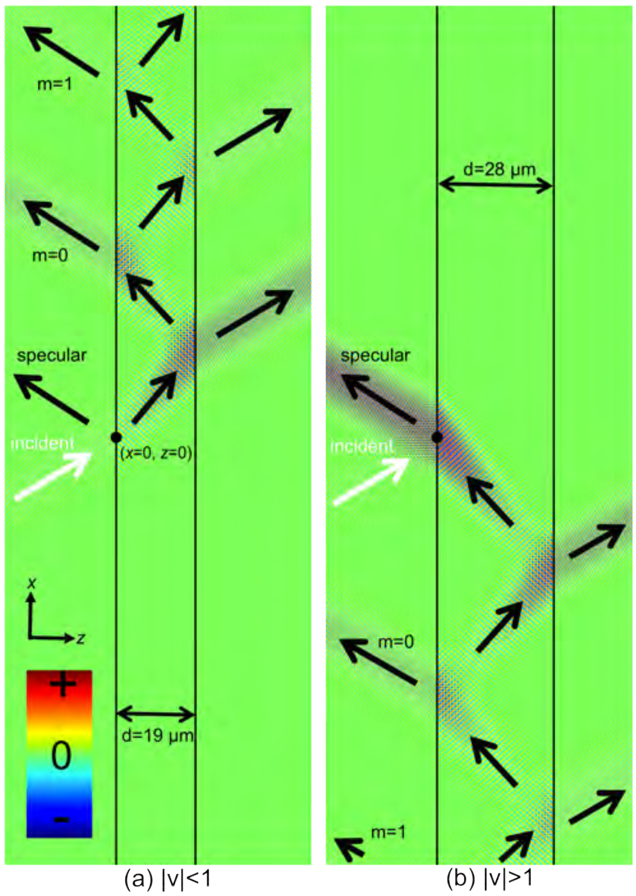

FIG. 2. Plots of the field $E_y(x,z)$ at one instant of time for a Gaussian beam (indicated with the white arrow) incident on an amplifying slab for which (a) $|\nu| < 1$ ($d = 19$ μm) and (b) $|\nu| > 1$ ($d = 28$ μm). The black dot indicates the origin of the coordinate system. For the same material parameters as (b), an incident pulse of light more vividly illustrates the peculiar behavior (Multimedia view).

In Fig. 2(b) all parameters are kept the same except the slab thickness is increased to $d = 28$ μm, resulting in $|\nu| > 1$ for all constituent planewaves of the Gaussian beam ($1.01 < |\nu| < 2.58$). Based solely on the plot of the field amplitude and not on the direction of energy flow indicated by the arrows, it may appear that the incident beam strikes the interface and negatively refracts in the slab, then zig-zags downwards in the $-\hat{\boldsymbol{x}}$ direction, giving rise to many reflected beams in medium one (and transmitted beams in medium three) which emanate from points on the slab with $x < 0$. Such an explanation was offered for simulations similar to ours [14, 15] to attempt to justify negative refraction in an active, nonmagnetic medium. However, by analyzing the Poynting vector we see that the energy in the beam zig-zags *up* the slab, so this phenomenon is distinct from negative refraction, despite the similarity in the positions of the reflected and transmitted beams. A video of a Gaussian pulse of light with a temporal FWHM of 50 fs and all other parameters identical to those of Fig. 2(b) more vividly illustrates that energy flows in the $+x$-direction (Multimedia view). The video plots the pulse intensity–not amplitude–on a logarithmic scale covering three decades. We will refer to the field in the slab at $x < 0$ as the "pre-excitation," so-called because it occurs before the central lobe of the incident beam arrives at the slab. Each reflected beam in Fig. 2(b) can be associated either with the specular term r'_{12} or with the mth term of the primed partial wave expansion in Eq. 7.

The Fresnel solution for a slab with $|\nu| > 1$ is a steady-state harmonic solution, in the sense that if the field distribution presented in Fig. 2(b) exists at time t_o, then as time is evolved forward the field at each point in space will vary harmonically with frequency ω. The intent of this Letter is not to investigate the causal evolution of the pre-excitation, beginning with the time the excitation source is turned on. We also recognize that the experimental verification of this potential phenomenon will be complicated by factors not included in the Fresnel formalism, such as spontaneous emission, that could lead to instabilities or self-lasing. Experimental work already done on the amplification of evanescent waves [16, 23, 24] (a regime for which $|\nu| > 1$), however, has not suffered from either of these problems. We note also that the Fresnel formalism, by implicitly beginning with the time-harmonic subset of Maxwell's equations, can only elucidate the non-divergent solutions to the full time-dependent equations. There can certainly exist divergent solutions, as demonstrated by finite-element simulations of a wave with a well-defined start-time incident normally on a slab with $|\nu| > 1$ [6].

Rather, we take the pre-excitation behavior demonstrated in Fig. 2(b) as the direct, logical, and inescapable consequence of the Fresnel formalism applied to Maxwell's equations for situations in which $|\nu| > 1$. This solution has surfaced in the literature [1–21], sometimes knowingly but often not, but its peculiar properties have not been sufficiently appreciated. Our intent is merely to explore these properties, and explain their relevance to some persistent controversies.

We observe that the specular reflection is given by r_{12} when $|\nu| < 1$ and r'_{12} when $|\nu| > 1$. Since $|r_{12}| < 1$ (in most cases of practical interest) and $r'_{12} = 1/r_{12}$, this means that $|r'_{12}| > 1$, and so the primed partial wave expansion mathematically predicts the amplification of the specularly reflected beam when $|\nu| > 1$. From Fig. 2(b) we see that this amplification occurs because the specular beam receives energy from the transmission of the pre-excited field through the 2-1 interface. Another noteworthy feature of this solution is that when $|\nu| \gg 1$ (achieved either by increasing the thickness or gain of the slab, or the incidence angle), the left-propagating wave amplitude E_{2L} becomes much larger than E_{2R}. This dominance of the left-propagating wave is of course a direct (although certainly peculiar) result of the multiple reflections of the pre-excitation at the front and back facets of the slab [3], without which only the right-propagating wave E_{2R}

would exist in medium two.

It turns out that TIR from an amplifying slab is well within the regime $|\nu| \gg 1$ (for any reasonable thickness d). As θ surpasses the critical angle for TIR, $|\nu|$ quickly becomes extremely large due to the negatively increasing $\text{Im}(k_{2z}^R)$. (For the parameters used in Fig. 2(b), the critical angle is $\theta_c = 41.8°$. For $\theta = 41°$, $|\nu| = 9.34 \times 10^3$, and for $\theta = 42°$, $|\nu| = 1.40 \times 10^{15}$.) Therefore, only the left-propagating wave E_{2L} exists with any appreciable amplitude in the slab, and we emphasize again that this results directly from the multiple reflections of the pre-excitation at both slab facets. Some have argued that even if medium two is semi-infinite, above the critical angle the incident wave excites the wavevector k_{2z}^L in the transmission medium rather than the usual k_{2z}^R, resulting in the reflection coefficient r'_{21} and an accompanying amplified specular reflection [16–19, 21]. It seems to us, however, that since the existing experimental results [16, 24] can be explained by the slab picture without recourse to the single-surface problem, there is no need for a special TIR-exception to the postulate that k_{2z}^R is the transmitted wavevector in the single-surface problem.

In the case of the negative-index lens [1] (where media one and three are vacuum and medium two has ϵ_2=μ_2=-1), every incident evanescent wave ($k_x > \omega/c$), for both s and p-polarization, excites a lossless surface plasmon polariton mode [4] on the 1-2 and 2-3 interfaces, resulting in $r_{21} = r_{23} = \infty$ and hence $\nu = \infty$ (despite being a passive medium). Therefore, the convergent solution found by applying the Fresnel formalism to this problem [1, 2] is the one described by the primed geometric series in Eq. 7 with $\nu' = 1/\nu = 0$. The result is that only the wavevector k_{2z}^L exists in the slab, which describes an evanescent wave that is amplified with increasing z, and therefore enables the "perfect lensing" action. In this case, the reflection coefficient from the slab is given by $r = r'_{12} = 0$, and our argument indicates that this is the result of multiple reflections within the slab [3], *not* because the incident evanescent wave is impedance-matched to the slab [2, 5]. If loss is introduced to the slab, r_{21} and r_{23} become finite but the lens still works well if $|\nu| \gg 1$; however, even small losses lead to $|\nu| < 1$, causing the decaying wave k_{2z}^R to dominate the amplified wave k_{2z}^L, thereby spoiling the perfect lens [4, 5].

In conclusion, we have shown that the convergence of the Fresnel solution for a cavity with net roundtrip gain relies on the existence of the 'pre-excited' field, which is a peculiar manifestation of the geometric partial wave series. By elucidating this counterintuitive phenomenon, we hope to have provided a useful alternative perspective for understanding amplified total internal reflection and the negative-index lens. We have also shown a positive-index slab with net roundtrip gain does not negatively refract–however, because the behavior mimics negative refraction insofar as the positions of the reflected and transmitted beams are concerned, the slab could substitute as a negative-index material in certain applications.

The pulse simulation was run on the Odyssey cluster supported by the Harvard FAS Research Computing Group. TSM is supported by an NSF Graduate Research Fellowship. We thank Alexey Belyanin and Steven Byrnes for helpful discussions.

[1] J.B. Pendry. Phys. Rev. Lett. **85**, 3966 (2000).
[2] G.W. 't Hooft. Phys. Rev. Lett. **87**, 249701 (2001).
[3] J.B. Pendry. Phys. Rev. Lett. **87**, 249702 (2001).
[4] D.R. Smith, D. Schuring, M. Rosenbluth, S. Schultz, S.A. Ramakrishna, and J.B. Pendry. Appl. Phys. Lett. **82**, 1506 (2003).
[5] N. Garcia and M. Nieto-Vesperinas. Phys. Rev Lett. **88**, 20, 207403 (2002).
[6] X. Jiang, Q. Li and C.M. Soukoulis. Phys. Rev. B **19**, 9007 (1999).
[7] S.A. Ramakrishna. Pramana **62**, 1273, (2004).
[8] J.C.J. Paasschens, T. Sh. Misirpashaev and C.W.J. Beenakker. Phys. Rev. B **54**, 11887 (1996).
[9] P. Pradhan and N. Kumar. Phys. Rev. B **50**, 9644 (1994).
[10] Z.Q. Zhang. Phys. Rev. B **52**, 7960 (1995).
[11] Y. Chen, P. Fischer, and F.W. Wise. Phys. Rev. Lett. **95**, (2005).
[12] S.A. Ramakrishna. Phys. Rev. Lett. **98**, (2007).
[13] S.A. Ramakrishna and O.J.F. Martin. Opt. Lett. **30**, 19, 2626 (2005).
[14] J.B. Geddes III, T.G. Mackay, and A. Lakhtakia. Opt. Comm. **280**, 120 (2007).
[15] B. Nistad and J. Skaar. Phys. Rev. E **78**, 036603 (2008).
[16] C.J. Koester. IEEE J. Quantum Electron. **QE-2**, 580 (1966)
[17] P.R. Callary and C.K. Carniglia. J. Opt. Sci. Am. **66**, 8 (1976).
[18] J. Fan, A. Dogariu, and L.J. Wang. Opt. Express **11**, 299 (2003).
[19] K.J. Willis, J.B. Schneider, and S.C. Hagness. Opt. Express **16**, 1903 (2008).
[20] A. Siegman. OPN, 38 (Jan. 2010).
[21] J.O. Grepstad and J. Skaar. Opt. Express **19**, 21404 (2011).
[22] See supplemental material at [URL will be inserted by AIP] for details on the R/L superscript convention, a discussion of lasing, description of the pulse video, and details of the simulation method.
[23] E.P. Ippen and C.V. Shank. Appl. Phys. Lett. **21**, 301 (1972).
[24] H.H. Choi, H.J. Kim, J. Noh and W. Jhe. Jour. Korean Phys. Soc. **42**, 80 (2003).