Lasers with sublinear output power characteristic due to nonuniform gain saturation

Tobias S. Mansuripur,[1] Guy-Mael de Naurois,[2] Alexey Belyanin,[3] and Federico Capasso[2,*]

[1]Department of Physics, Harvard University, Cambridge, MA 02138 USA
[2]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA
[3]Department of Physics and Astronomy, Texas A & M University, College Station, Texas 77843 USA

It is a generally accepted fact of laser physics that in a homogeneously broadened gain medium, above threshold the output power of the laser grows linearly with the pump power. The derivation requires only a few simple lines in laser textbooks, and the linear growth is a direct result of the fact that above threshold, the intracavity optical intensity will increase to the point that the gain is saturated to the level of the net loss–so-called gain "pinning" or "clamping." Such a derivation, however, assumes that the intensity is constant across the length of the gain medium–the approximation of uniform gain saturation–which is only a good assumption for cavities whose end mirrors have reflectivities close to one. We prove that for cavities with large out-coupling that also have significant distributed loss, above threshold the output power at first grows sublinearly with the pump power. At high pump power the output grows linearly, but with a slope efficiency that can be substantially smaller than the prediction of the uniform gain saturation theory, with the largest deviation occurring for traveling-wave lasers and asymmetric Fabry-Pérot lasers. These results are particularly applicable to semiconductor lasers, and specific applications to quantum cascade lasers are discussed.

I. INTRODUCTION

One of the signature characteristics of laser action in homogeneously broadened media is often proclaimed to be the linear growth of the output power with the pump power above the lasing threshold. In textbook derivations [1–7] this behavior is seen to be the direct result of gain clamping: above threshold, a steady state can only be reached when the intracavity optical intensity is such that the rate of stimulated emission reduces the population inversion to its threshold value, thereby always pinning the gain to the level of the net loss. Such a derivation, however, assumes that the gain saturates uniformly–in other words, that the intensity is constant throughout the gain medium–which is only a good assumption for cavities whose end mirrors have reflectivities close to one. We show that this approximation can lead to unphysical predictions for the output power in cavities with large out-coupling that also contain a distributed intrinsic loss; in such cavities, there is a maximum sustainable intracavity intensity that is proportional to the pump power at which the gain due to stimulated emission is exactly compensated by the intrinsic loss [8–10]. In cases where the intracavity intensity approaches this maximum, the uniform gain saturation treatment of the laser will overestimate the output power. We will apply the equations first popularized by Rigrod [9, 11–15] to account for the nonuniformity of the gain saturation and the lumped mirror losses, and demonstrate that–in the presence of distributed loss–the output power grows sublinearly with the pump power after threshold. Far above threshold the output grows linearly, but with a slope efficiency that can be significantly less than one would expect from the uniform gain saturation theory. The deviation of the output power from the prediction of the uniform gain saturation theory is largest for traveling-wave lasers–which display the largest variation in intracavity intensity–and also very significant for highly asymmetric Fabry-Pérot (FP) cavities (i.e., one mirror with reflectivity much larger than the other). In symmetric FP cavities the deviation is the smallest, but can still be significant for particularly large values of distributed loss and out-coupling.

In what follows, we begin by explaining the geometry of the ring and FP cavities under consideration and introduce the Rigrod equation for homogeneously broadened gain media. A physical picture is presented to understand the uniform gain approximation in which the lumped transmission losses at the mirrors are replaced with a distributed out-coupling loss. With this picture in mind, it is simple to understand why the uniform gain approximation must fail for lasers with distributed intrinsic loss. Taking into account nonuniform gain saturation, we first analyze the traveling-wave laser, which yields an explicit formula for the output power. We then explore a few specific examples of the FP cavity using numerical solutions, and use the results to draw general conclusions about the behavior of both symmetric and asymmetric FP cavities. Finally, we point out how these new results are used to optimize laser design, using quantum cascade lasers (QCLs) as a model system.

A simple FP laser cavity is shown in Fig. 1(a), which comprises a gain medium of length L placed between two lossless mirrors with reflectivities R_1 and R_2 (and transmissivities $1 - R_1$ and $1 - R_2$). (The facets of the gain medium itself do not reflect, either through the use of Brewster-cut facets or anti-reflection coatings.) The intensity envelopes of the right and left-propagating waves are $I^+(z)$ and $I^-(z)$. Following the notation of Rigrod [12], we will use the dimensionless ra-

* capasso@seas.harvard.edu

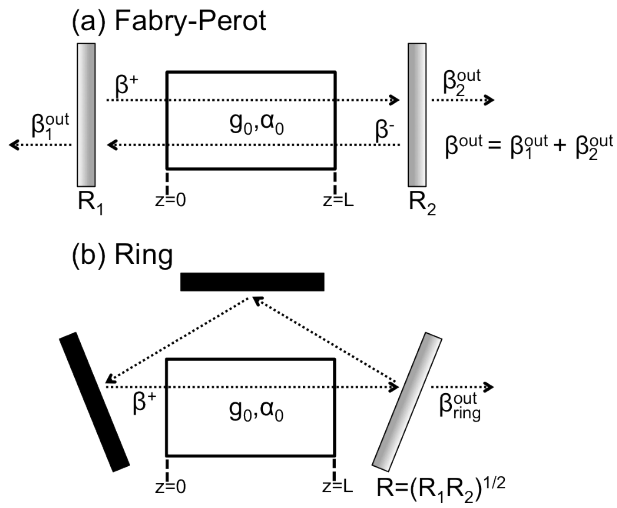

FIG. 1. The laser cavity geometry and relevant parameters for both a (a) FP cavity and (b) unidirectional traveling-wave ring cavity.

tios $\beta^+(z) = I^+(z)/I_{\rm sat}$ and $\beta^-(z) = I^-(z)/I_{\rm sat}$, where $I_{\rm sat}$ is the saturation intensity of the medium. The unsaturated gain, which is typically proportional to the pump power, is g_0 per unit length and the distributed intrinsic loss is α_0 per unit length. In Fig. 1(b), the same gain medium is employed in a traveling-wave configuration by placing it within a ring cavity comprising two perfect reflectors and one out-coupling mirror of reflectivity R. We will consider the unidirectional mode for which only β^+ is nonzero (achieved in practice with an optical isolator).

For a homogeneously broadened gain medium, the intensity envelopes obey the equations [1, 12–16]

$$\frac{1}{\beta^+}\frac{d\beta^+}{dz} = -\frac{1}{\beta^-}\frac{d\beta^-}{dz} = \frac{g_0}{1+\beta^+ +\beta^-} - \alpha_0. \qquad (1)$$

We will consider only the homogeneously broadened case here; special cases of inhomogeneous broadening are discussed in [9, 16, 17]. By dealing only with the intensity envelope of the field, population grating effects resulting from spatial hole burning (SHB) are not taken into account, but this treatment is nevertheless known to yield results in good agreement with experiment [1]. Equation 1 applies strictly to a single transverse and longitudinal mode; in a multimode laser the shape of the gain spectrum becomes important as does the mutual cross-saturation of the various frequency components. While SHB is known to lead to multimode operation in homogeneously broadened lasers [16], SHB is not as significant in ring and asymmetric FP lasers, which display predominantly traveling-wave rather than standing-wave character. It is worth noting that our main conclusion in this paper is most relevant to these two types of cavities where SHB is less important. Moreover, we expect our results to still be valid in the multimode regime because the shape of the intensity envelope of each mode should not differ significantly from the single-mode case.

While Eq. 1 (together with appropriate boundary conditions) is needed to solve for the laser output power above threshold, exactly at threshold the gain is completely unsaturated (i.e., intensity is zero) so the threshold condition is derived simply by demanding that the roundtrip unsaturated gain compensates both the intrinsic and out-coupling losses of the cavity. (Note that this derivation applies only when spontaneous emission is very weak, which is a good assumption for many kinds of lasers. A treatment that includes gain saturation at threshold due to spontaneous emission is given in [18, 19].) To facilitate comparisons between the FP and ring cavities, we choose the reflectivity of the ring cavity out-coupler to be $R = \sqrt{R_1R_2}$ so that both cavities have the same threshold, given by

$$g_{0,th} = \alpha_0 + \frac{1}{L}\ln\left(\frac{1}{R}\right). \qquad (2)$$

II. DISTRIBUTED LOSS APPROXIMATION

The second term on the right-hand side of Eq. 2 is commonly called the mirror loss $\alpha_m = \ln(1/R)/L$, so that the total loss can be written as $\alpha_0 + \alpha_m$. We emphasize, however, that these two contributions to the loss behave very differently, α_0 being a distributed loss and α_m a lumped loss. A significant approximation made in textbook derivations of the laser output power [6, 7]–though not always explicitly stated–is to treat α_m as a distributed loss. In this way, all losses of the cavity become distributed, so we will refer to this as the distributed loss approximation (DLA). One way to think of the DLA is to imagine that the end mirrors of the cavity are replaced by perfect reflectors, and light escapes instead from any point within the gain medium with decay constant α_m. (Equivalently, a photon escape time $\tau_{\rm esc}$ is often defined as $\tau_{\rm esc}^{-1} \equiv \alpha_m v_g$, where v_g is the group velocity of light in the cavity.) Because there are no longer any lumped losses, in the steady state the intracavity intensity must be independent of position or else it would diverge. In other words, the saturated gain due to stimulated emission is exactly cancelled by the sum of the intrinsic loss and mirror loss at every position in the cavity, which is why this is usually called the uniform gain saturation approximation. Mathematically, Eq. 1 becomes

$$\frac{1}{\beta^+}\frac{d\beta^+}{dz} = \frac{g_0}{1+\beta} - (\alpha_0+\alpha_m) = 0, \qquad (3)$$

where $\beta \equiv \beta^+ + \beta^-$ is the total intensity. Equation 3 is easily solved for β and multiplied by the distributed out-coupling $\alpha_m L$ (as opposed to the lumped out-coupling $1-R$, which is not the relevant out-coupling in the DLA) to get the total out-coupled intensity

$$\beta^{\rm out}_{\rm DLA} = \frac{\alpha_m L}{\alpha_0+\alpha_m}\left(g_0 - g_{0,th}\right). \qquad (4)$$

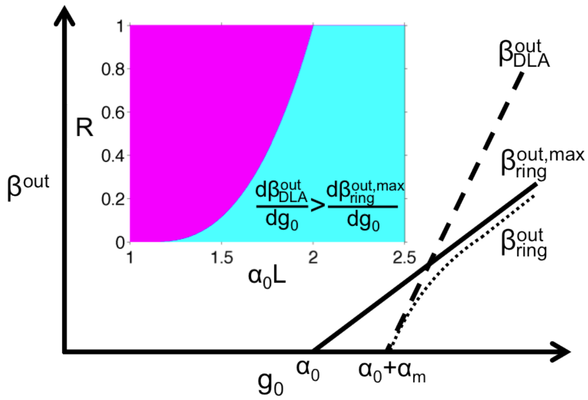

FIG. 2. (a) A sketch illustrates that if the slope of $\beta^{\rm out}_{\rm DLA}$ (dashed) exceeds the slope of $\beta^{\rm out,max}_{\rm ring}$ (solid), the DLA prediction exceeds the allowable maximum output of the ring cavity above a certain gain. Therefore, we can expect the actual output $\beta^{\rm out}_{\rm ring}$ to qualitatively follow the dotted curve. Inset: the parameter ranges of $\alpha_0 L$ and R that result in such a violation, namely, those which satisfy Inequality 6, are shaded in light blue.

In the DLA, both the FP and ring cavities emit the same intensity $\beta^{\rm out}_{\rm DLA}$. Note that the linearity of $\beta^{\rm out}_{\rm DLA}$ with the unsaturated gain g_0–and therefore, in most cases, the pump power–follows from the DLA almost immediately.

The parameter ranges for which the DLA is a good approximation will be discussed in Sec. III and Appendix A. For now, we point out one pitfall of the DLA that is central to our main result. In a gain medium with intrinsic loss, whether an amplifier or a laser, there is a maximum attainable intracavity intensity $\beta^{\rm max}$ at which the number of photons generated by stimulated emission is exactly equal to the number of photons dissipated by the intrinsic loss [8–10]. By setting Eq. 1 to zero, we find that the total intracavity intensity can never (in the absence of external light injection) exceed $\beta^{\rm max} = (g_0 - \alpha_0)/\alpha_0$. This, in turn, places an upper limit on the out-coupled intensity. For simplicity, we consider the ring cavity to explain the basic point: the output cannot exceed the mirror transmission times $\beta^{\rm max}$,

$$\beta^{\rm out,max}_{\rm ring} = (1-R)\beta^{\rm max} = \frac{1-R}{\alpha_0}(g_0 - \alpha_0). \qquad (5)$$

(Similar considerations apply to the FP cavity and will be discussed in Sec. III B.) It turns out that the DLA, by virtue of replacing the lumped mirror transmission $1-R$ with the distributed out-coupling $\alpha_m L$, yields predictions for the output intensity that violate this maximum under certain conditions. Specifically, whenever the slope efficiency of $\beta^{\rm out}_{\rm DLA}$ is greater than the slope efficiency of $\beta^{\rm out,max}_{\rm ring}$,

$$\frac{\alpha_m L}{\alpha_0 + \alpha_m} > \frac{1-R}{\alpha_0}, \qquad (6)$$

the sketch in Fig. 2 illustrates that $\beta^{\rm out}_{\rm DLA}$ will exceed $\beta^{\rm out,max}_{\rm ring}$ above some value for g_0. In the next section we will solve explicitly for the output $\beta^{\rm out}_{\rm ring}$ taking into account nonuniform gain saturation, but we can already guess that the curve should resemble the dotted line in Fig. 2: it has the same threshold as the DLA predicts, but the output power must eventually be limited by $\beta^{\rm out,max}_{\rm ring}$. The parameter ranges of $\alpha_0 L$ and R that satisfy Inequality 6 are shown in light blue in the inset; in general, as $\alpha_0 L$ becomes larger and R smaller, the limiting influence of $\beta^{\rm out,max}_{\rm ring}$ becomes more relevant and nonuniform gain saturation must be accounted for. (In the purple region of the inset, there is only a small difference between the predictions for the output intensity as given by the DLA and the nonuniform gain saturation approach.)

III. NONUNIFORM GAIN SATURATION

We now address the behavior of the laser when the lumped nature of the mirror losses is properly accounted for, and the gain saturates nonuniformly within the cavity. The general solution of Eq. 1 is given in [12], which we will restate here. It can be deduced from Eq. 1 that the product $\beta^+(z)\beta^-(z)$ is independent of z, so we designate the quantity $\beta_0^2 \equiv \beta^+(z)\beta^-(z)$. By making the substitution $\beta^-(z) = \beta_0^2/\beta^+(z)$, Eq. 1 is expressed entirely in terms of $\beta^+(z)$, which can then be integrated to yield

$$\alpha_0 L - \ln(\beta_1^+/\beta_2^+) = \frac{g_0 \ln[F(\beta_1^+)/F(\beta_2^+)]}{\sqrt{(g_0-\alpha_0)^2 - (2\alpha_0\beta_0)^2}}, \qquad (7)$$

where

$$F(\beta_i^+) \equiv \frac{\sqrt{(g_0-\alpha_0)^2 - (2\alpha_0\beta_0)^2} + (g_0 - \alpha_0 - 2\alpha_0\beta_i^+)}{\sqrt{(g_0-\alpha_0)^2 - (2\alpha_0\beta_0)^2} - (g_0 - \alpha_0 - 2\alpha_0\beta_i^+)}. \qquad (8)$$

The subscript i on β^+ is either 1 or 2, and denotes evaluation of the intensity at the left facet ($z=0$) or right facet ($z=L$), respectively. Note that Eqs. 7 and 8 apply to both the FP and ring cavity. For the FP cavity, we use the boundary conditions at the facets

$$\beta_1^+ = R_1\beta_1^-, \; \beta_2^- = R_2\beta_2^+ \qquad (9)$$

together with

$$\beta_0^2 = \beta_1^+\beta_1^- = \beta_2^+\beta_2^- \qquad (10)$$

to show that $\beta_1^+ = R\beta_2^+$ and $\beta_0 = \sqrt{R_2}\beta_2^+$, which allows us to express Eq. 7 entirely in terms of β_2^+. Still, in the general FP case Eq. 7 remains an implicit formula for β_2^+ which must be solved numerically. For the ring cavity we substitute $\beta_1^+ = R\beta_2^+$ and $\beta_0 = 0$ into Eq. 7, which yields an analytic solution that will be discussed shortly.

In Appendix A, we discuss two limiting cases for which the total output intensity as predicted by the nonuniform gain saturation theory in Eq. 7 reduces to that of the

DLA: 1) small cavity out-coupling and arbitrary intrinsic loss α_0 and 2) arbitrary out-coupling and $\alpha_0 = 0$ (a surprising result, since the intensity variation in the cavity can be very large). In what follows, we will examine the case of large out-coupling and large intrinsic loss, for which the predictions of the nonuniform gain saturation theory can deviate significantly from those of the DLA.

A. Ring cavity

Unlike the case of the FP cavity, the equations for the unidirectional ring cavity can be solved for an explicit expression of the output power. (The existence of this solution–but not its expression–was stated in [12], and left unexplored because it did not further the goal of that paper, which was to derive an explicit expression for the optimal out-coupling.) Starting from Eqs. 7 and 8 and using $\beta_0 = 0$ and $\beta_1^+ = R\beta_2^+$, we find

$$\beta_{\rm ring}^{\rm out} = (1-R)\,(g_0/\alpha_0 - 1) \times \frac{R\exp[(1-\alpha_0/g_0)(\alpha_0+\alpha_m)L] - 1}{R\exp[(1-\alpha_0/g_0)(\alpha_0+\alpha_m)L] - R}. \quad (11)$$

We will explore various limits of Eq. 11. Close to threshold, the slope efficiency can be found by Taylor expanding the derivative of Eq. 11 with respect to g_0 to first order in the parameter $(g_0 - g_{0,th})/(\alpha_0 + \alpha_m)$, which gives

$$\frac{d\beta_{\rm ring}^{\rm out}}{dg_0} = \frac{\alpha_m L}{\alpha_0+\alpha_m} - \frac{\alpha_0 L}{(\alpha_0+\alpha_m)^2}\left(\frac{1+R}{1-R}\ln(1/R) - 2\right)(g_0 - g_{0,th}). \quad (12)$$

(We have also only retained terms to first order in $\alpha_0 L$ for simplicity, though the following conclusions hold for large $\alpha_0 L$.) At threshold (i.e., $g_0 = g_{0,th}$), the slope efficiency is given by the first term of Eq. 12, which is the same as the prediction of the DLA in Eq. 4. However, while the DLA predicts a constant slope efficiency above threshold, in fact the slope efficiency decreases as g_0 increases past $g_{0,th}$ as given by the second term of Eq. 12. (Note that the term in large parentheses is always positive for $0 < R < 1$.)

As g_0 continues to increase, the slope efficiency asymptotically approaches a constant value, which we can see by Taylor expanding Eq. 11 in the high-gain limit $g_0 \gg g_{0,th}$ (without making any assumption about $\alpha_0 L$), which gives

$$\beta_{\rm ring}^{\rm out}\big|_{g\gg g_{0,th}} = \frac{(1-R)(1-e^{-\alpha_0 L})}{\alpha_0(1-Re^{-\alpha_0 L})}(g_0 - g_{0,th}^{\rm eff}) \quad (13)$$

where

$$g_{0,th}^{\rm eff} \equiv (\alpha_0+\alpha_m)(\alpha_0 L)\left(\frac{1}{1-e^{-\alpha_0 L}} - \frac{1}{1-Re^{-\alpha_0 L}}\right) + \alpha_0. \quad (14)$$

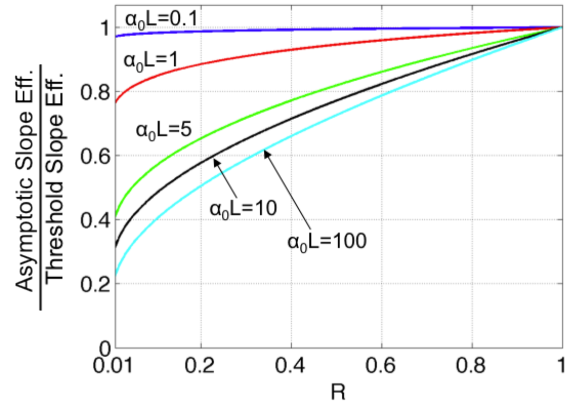

FIG. 3. The ratio of the asymptotic slope efficiency of the ring cavity to its threshold slope efficiency as a function of R for various values of $\alpha_0 L$.

We will refer to the derivative of Eq. 13 with respect to g_0 as the "asymptotic slope efficiency," which is clearly independent of g_0. In the limit of "moderate to large" intrinsic loss, $\exp(\alpha_0 L) \gg 1$, Eq. 13 reduces to

$$\beta_{\rm ring}^{\rm out}\big|_{g\gg g_{0,th}\ \&\ \exp(\alpha_0 L)\gg 1} = \frac{1-R}{\alpha_0}(g_0 - \alpha_0), \quad (15)$$

which is recognizable as $\beta_{\rm ring}^{\rm out,max}$ of Eq. 5. Thus, we see that for moderate to large $\alpha_0 L$, at high gain the intracavity intensity reaches its maximum $\beta^{\rm max}$, so the output intensity follows $\beta_{\rm ring}^{\rm out,max}$, as illustrated by the dotted line in Fig. 2. For small $\alpha_0 L$, the output intensity requires the full Eq. 13; in this case $\beta^{\rm max}$ is very large and the intracavity intensity never grows sufficiently to approach $\beta^{\rm max}$.

To understand quantitatively how much the asymptotic slope efficiency differs from the near-threshold slope efficiency, let us look at the ratio of the slope efficiency in the high-gain limit (slope of Eq. 13) to the slope efficiency at threshold (the first term of Eq. 12, which is the same as the slope efficiency predicted by the DLA). This ratio is plotted in Fig. 3 as a function of R for five different values of the product $\alpha_0 L$. As expected, the discrepancy between the asymptotic and threshold slope efficiency is greatest for large $\alpha_0 L$ and small R. This can be understood in terms of Fig. 2: as α_0 increases the slope of $\beta_{\rm ring}^{\rm out,max}$ decreases, and as L increases the slope of $\beta_{\rm DLA}^{\rm out}$ increases. In either case, an increase in $\alpha_0 L$ exaggerates the difference between the asymptotic and threshold slope efficiencies.

Finally, while the asymptotic output intensity was derived in the limit $g_0 \gg g_{0,th}$, it turns out that this is not a very stringent limit. For a full quantitative analysis, see Appendix B. Here we emphasize the conclusion of that analysis: the full solution in Eq. 11 approaches the asymptotic solution in Eq. 13 very closely when g_0 is only a few multiples of $g_{0,th}$, and often when g_0 is even smaller. This is significant because in some lasers, such as QCLs,

g_0 can never exceed more than a few multiples of $g_{0,th}$, and so this asymptotic regime is still experimentally accessible. As $\alpha_0 L$ is increased at fixed R, the asymptotic regime is reached at ever smaller values of g_0. This trend can be understood easily with recourse to Fig. 2: note that the gain-intercept of $\beta_{\text{ring}}^{\text{out,max}}$ is α_0 while the laser threshold is $\alpha_0 + \alpha_m$. As $\alpha_0 L$ increases at fixed R, the intercept α_0 approaches $\alpha_0 + \alpha_m$, and so the laser output must approach $\beta_{\text{ring}}^{\text{out,max}}$ sooner after threshold. In other words, the asymptotic regime is reached soon after threshold. Let us consider one numerical example: for $\alpha_0 L = 10$ and $R = 0.25$, the asymptotic slope efficiency is 62% of the threshold slope efficiency (see Fig. 3), and the asymptotic regime is reached once $g_0 \geq 1.4 g_{0,th}$ (see Appendix B). This example will be extended to the FP cavity in the next section.

B. FP cavity

The FP cavity does not yield an explicit expression for the output power, but the problem can of course be solved numerically. As a concrete example, we consider the case of $\alpha_0 = 20$ cm^{-1}, $L = 0.5$ cm, and $R = 0.25$. (These are representative values for a QCL with emission wavelength between 8 and 12 μm [20].) In Fig. 4, the total output power β^{out} as a function of g_0 is shown for three lasers: 1) symmetric FP with $R_1 = R_2 = 0.25$, 2) maximally asymmetric FP with $R_1 = 1$, $R_2 = 0.25^2 = 0.0625$, and 3) ring cavity configuration with $R = 0.25$. According to the DLA, all three lasers should have the same output power, and this curve is plotted as well. The slope efficiency is shown in the inset. We see that all three lasers have the same slope efficiency at threshold, and the value is correctly predicted by the DLA. (For the ring cavity we already proved this fact in the previous section.) As the gain increases, the slope efficiency of all three lasers decreases. Once the asymptotic regime is reached, the symmetric FP has the largest slope efficiency (96% of threshold value), followed by the asymmetric FP (72% of threshold value) and then the ring cavity (62% of threshold value). Note that it should be possible to achieve any value for the asymptotic slope efficiency between 72% and 96% of the threshold value for various choices of R_1 and R_2.

We can understand the reason for these differences in output power among the lasers by looking at Fig. 4(b), which plots the intracavity intensity at the right mirror (where it is a maximum), $\beta_2^+ + \beta_2^-$, normalized to β^{max}. In the ring cavity and asymmetric FP, shortly after threshold the total intensity at the right mirror quickly approaches β^{max}. In the symmetric FP, the total intensity at the right mirror asymptotically approaches the slightly smaller value of $0.973\beta^{\text{max}}$, although for larger values of $\alpha_0 L$ it can be shown that the intensity in the symmetric FP will also approach β^{max}. Thus, for sufficiently large $\alpha_0 L$, an accurate approximation is to say that the maximum intracavity intensity of all three cavities reaches β^{max} for large g_0. The output intensity of each cavity therefore approaches its theoretical maximum, which is easily shown to be

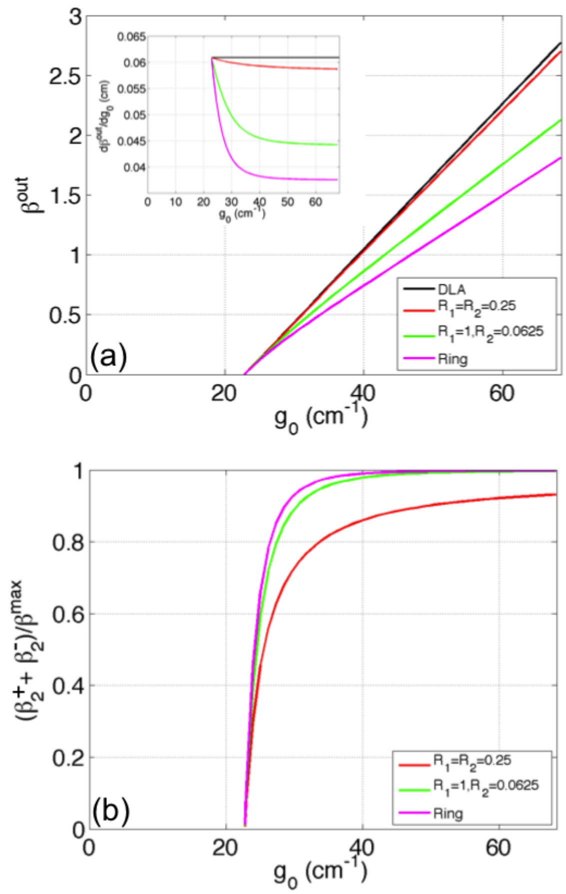

FIG. 4. (a) β^{out} vs. g_0 for $\alpha_0 = 20$ cm^{-1}, $L = 5$ mm, and $R = 0.25$, shown for the DLA (black), symmetric FP with $R_1 = R_2 = 0.25$ (red), maximally asymmetric FP with $R_1 = 1$ and $R_2 = 0.25^2 = 0.0625$ (green), and ring cavity configuration with $R = 0.25$ (magenta). The slope efficiency of the curves (i.e., first derivative) is shown in the inset, demonstrating that all configurations have the same slope efficiency at threshold, which then decreases with increasing g_0 and approaches an asymptotic value. (b) The intracavity intensity at the right facet, $\beta_2^+ + \beta_2^-$, is normalized to β^{max} and plotted against g_0. Note that β^{max} is itself an increasing function of g_0.

$$\beta_{\text{sym}}^{\text{out,max}} = 2\left(\frac{1-R}{1+R}\right)\beta^{\text{max}} \qquad (16)$$

$$\beta_{\text{asym}}^{\text{out,max}} = \left(\frac{1-R^2}{1+R^2}\right)\beta^{\text{max}} \qquad (17)$$

for the symmetric FP and maximally asymmetric FP, respectively, while the maximum ring laser output was already given in Eq. 5. (For the maximally asymmetric FP, the reflectivity of one facet is unity and the other is $R_2 = R^2$.) The denominators $1+R$ and $1+R^2$ in Eqs. 16 and 17, respectively, arise from the need to transform the total intracavity intensity at the right mirror, β^{max}, into

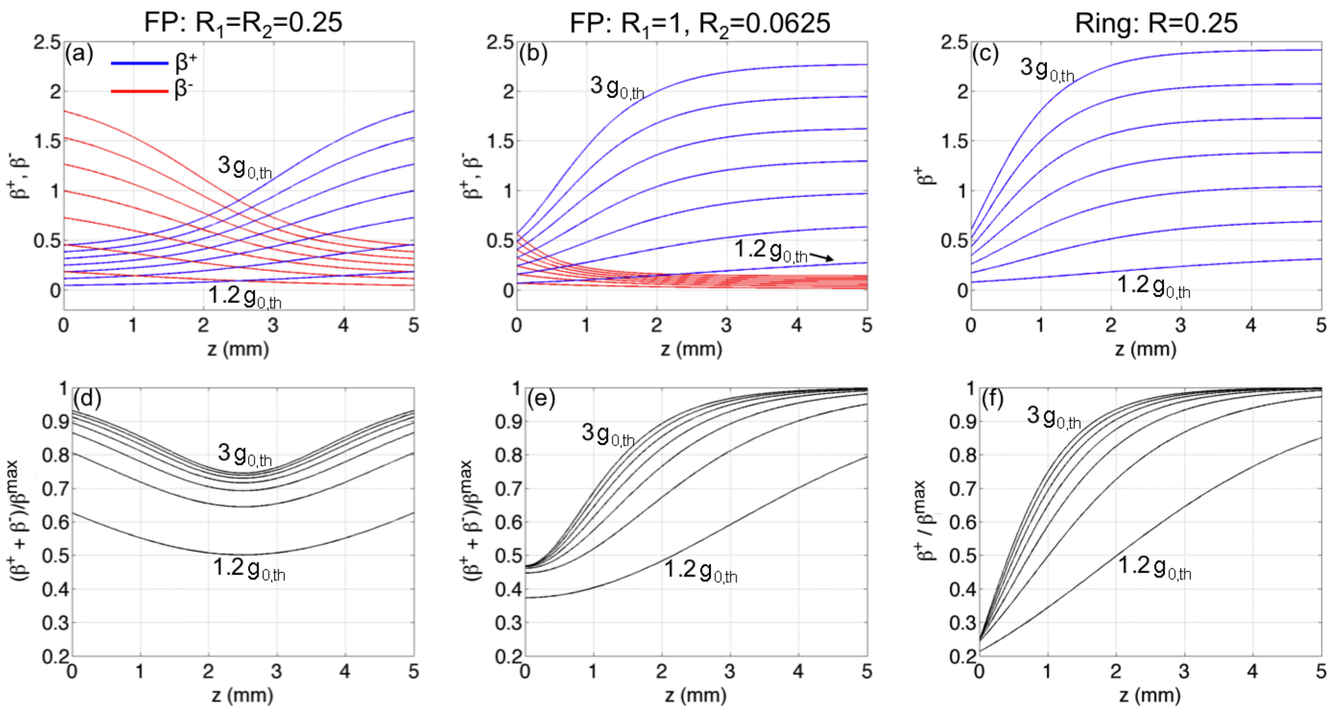

FIG. 5. For the same parameters as those used in Fig. 4 ($\alpha_0 = 20$ cm^{-1}, $L = 5$ mm, $R = 0.25$), the right and left-propagating intensity envelopes β^+ and β^- are plotted for $1.2g_{0,th}$ to $3g_{0,th}$ (in steps of $0.3g_{0,th}$) for a (a) symmetric FP with $R_1 = R_2 = 0.25$, (b) maximally asymmetric FP with $R_1 = 1$ and $R_2 = 0.25^2 = 0.0625$, and (c) ring cavity configuration with $R = 0.25$. The total intracavity intensity, $\beta^+ + \beta^-$, is normalized to $\beta^{\max}$ and plotted in (d),(e), and (f) for the same three cavities, respectively.

the right-traveling component β_2^+, which is then multiplied by the transmissivity $1 - R$ or $1 - R^2$, respectively, to arrive at the output intensity. Equations 5 and 16-17 can be useful when dealing with lasers with large intrinsic loss, but only when one has prior knowledge that the laser is operating within the asymptotic regime–in other words, that the intracavity intensity nears $\beta^{\max}$. For example, the formulas can be used to quickly approximate the relative output powers of the three lasers in Fig. 4(a) at high gain.

Though we have already shown that the intracavity intensity at the right mirror approaches $\beta^{\max}$ for all three cavities, a more complete understanding is obtained from plots of the intracavity intensity along the entire length of the cavity, as shown in Fig. 5 for various values of g_0. The symmetric FP has the most uniform intensity distribution as a result of the equal amplitudes of the left and right-propagating envelopes, and accordingly the uniform gain saturation approximation is the most accurate for this case. In the asymmetric FP, β^+ (which travels from the high-reflectivity mirror to the lower-reflectivity mirror) is much more intense than β^-. Therefore, β^- does not cross-saturate β^+ significantly, which allows β^+ of the asymmetric FP to reach a higher intensity than it does in the symmetric FP. As a result, β^+ can more closely approach $\beta^{\max}$ in the asymmetric FP, resulting in minimal amplification as it nears the right mirror. In the ring cavity, β^- is completely absent, thereby allowing β^+ to grow even more quickly at small z and approach the saturation limit $\beta^{\max}$ well before it reaches the right mirror. (See Appendix C for a mathematical discussion of the intensity variation in the various cavities.)

IV. OTHER APPLICATIONS

The DLA overestimates the increase in output power that can be achieved by increasing the length of the laser. As a practical example, we consider an asymmetric FP QCL with $\alpha_0 = 15$ cm^{-1}, high-reflectivity (HR) coating on one facet $R_1 = 1$, and low-reflectivity (LR) coating on the other $R_2 = 0.01$ (an achievable value with current technology) to increase the single-ended output power. In practice every laser has a maximum achievable gain g_0, so suppose our goal is to maximize the output power at $g_0 = 40$ cm^{-1}. In Fig. 6, the output power β^{out} vs. g_0 is plotted for two such lasers, one with $L = 4$ mm and one with $L = 5$ mm, for both the DLA and nonuniform gain saturation theory. In the DLA, the longer laser outputs 12% more power at $g_0 = 40$ cm^{-1} ($\beta^{\text{out}} = 2.40$ instead of 2.13). In reality, however, the longer laser can be expected to output only 3% more power ($\beta^{\text{out}} = 1.59$ instead of 1.54). The reason for this small increase is that the 4-mm laser is already emitting close to the maximum possible intensity $\beta^{\text{out,max}}_{\text{asym}}$ of Eq. 17 (also plotted in Fig. 6), so increasing the length of the laser will not

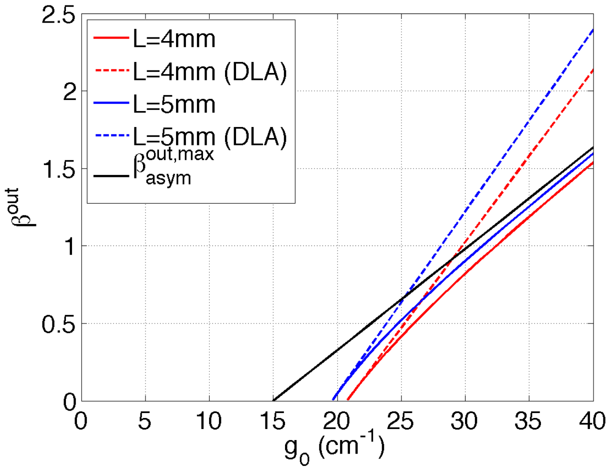

FIG. 6. The output power of the maximally asymmetric FP laser with $\alpha_0 = 15\ \text{cm}^{-1}$, $R_1 = 1$, and $R_2 = 0.01$, for cavities of length 4 mm (red) and 5 mm (blue), as computed numerically accounting for nonuniform gain saturation (solid line) and as predicted by the DLA (dashed line). The maximum output power $\beta^{\text{out,max}}_{\text{asym}}$ is also plotted (black).

increase the output significantly. In this example, a 25% increase in the length of the cavity results in only a 3% increase in power output, which clearly leads to a significant decrease in the wall-plug efficiency. More generally, for lasers with large intrinsic loss there is an optimum length beyond which the wall-plug efficiency will decrease.

The output intensity of a laser is maximized for a particular end mirror reflectivity, and this optimal reflectivity will be significantly miscalculated when using the DLA. (Note: Rigrod provides analytic formulas that account for nonuniform gain saturation to approximate the optimal reflectivity [12], but the formulas become less accurate for lasers with large intracavity intensity variation. This is the case for asymmetric FP and ring cavities, so here we will calculate the optimum numerically.) As before, let us consider $\alpha_0 = 15\ \text{cm}^{-1}$, $R_1 = 1$, and $L = 4$ mm, though now our goal is to find the optimum reflectivity R_2 that maximizes the power output at $g_0 = 40\ \text{cm}^{-1}$. The nonuniform gain saturation theory predicts an optimum reflectivity $R_2 = 5.4 \times 10^{-3}$ leading to $\beta^{\text{out}} = 1.54$. The DLA predicts an optimum reflectivity $R_2 = 5.0 \times 10^{-4}$, nearly an order of magnitude smaller. Using a coating with the DLA optimum reflectivity would result in $\beta^{\text{out}} = 1.50$, which is only a 2.6% reduction in output power relative to the true optimum. This is a small difference, but the importance of this example is that the optimal coating reflectivity is an order of magnitude larger than predicted by the DLA. This is important for active research in LR mid-infrared coatings; we see that we do not need to aim for reflectivities as low as we had thought, at least for the optimization of laser output power.

One final application concerns laser characterization. The ratio of the experimentally measured slope efficiencies of two lasers–identical except for their facet coatings, whose reflectivities are known–can be used to determine the value of α_0 by comparison with the DLA slope efficiency formula [21]. In such cases, the result can be skewed if one of the lasers is highly asymmetric, and a preferable method to determine the distributed loss would be to vary the cavity length L while maintaining the cavity symmetry $R_1 = R_2$ (or use the threshold values rather than the slope efficiencies, which are unaffected by the degree of symmetry of the cavity). (Note: the results of [21] are not affected by this because in that case one of the lasers had a HR-facet and a bare facet, as opposed to a LR-facet. Based on numerical calculations we have done, this would not have been enough asymmetry to affect their results by an experimentally significant margin.)

V. CONCLUSION

The uniform gain saturation treatment of a laser effectively replaces the transmission of light at the end mirrors with an out-coupling mechanism distributed throughout the cavity. We referred to this as the distributed loss approximation (DLA), and showed that it is a good approximation to the nonuniform gain saturation treatment when the out-coupling is small or when the intrinsic loss of the gain medium is zero. When the intrinsic loss is not zero, there is a maximum attainable intracavity intensity–proportional to the pump power–at which the number of photons generated by stimulated emission is equal to the number of photons absorbed or scattered by the intrinsic loss. When the intensity in the laser nears this maximum, the DLA overestimates the power that can be coupled out of the cavity. The result is that the output power grows sublinearly with the pump power, in stark contrast to the famous linear relationship derived in the uniform gain saturation approximation. The deviation from linear behavior is greatest for traveling-wave lasers and highly asymmetric FP cavities. Quantum cascade lasers, with lengths on the order of 1-5 mm and intrinsic loss between 5 and 20 cm^{-1}, have a large enough $\alpha_0 L$ product for this effect to be significant. Furthermore, it is common practice to maximize the output power from a single facet by using a HR-coating on one facet and a LR-coating on the other, thus enhancing the effect by increasing the asymmetry. It is important to understand the sublinear power output to optimally design and characterize lasers with large intrinsic loss.

FUNDING INFORMATION

We acknowledge support from the National Science Foundation under award number ECCS-1230477.

Appendix A: Limiting Cases of Nonuniform Gain Saturation Treatment

1. $R_1, R_2 \to 1$, $1 - R \ll \alpha_0 L$

When the reflectivity of both mirrors approaches one and there is also non-zero distributed loss (so that $1 - R \ll \alpha_0 L$), Eqs. 7 and 8 can be Taylor expanded to first order in the parameters $1 - R_1$ and $1 - R_2$, which yields the total output power

$$\beta^{\text{out}} \approx \frac{1 - R}{\alpha_0}(g_0 - \alpha_0). \tag{A1}$$

The DLA yields the same result as Eq. A1 when $\beta^{\text{out}}_{\text{DLA}}$ in Eq. 4 is expanded in the same limits. The agreement of the two theories in this limit is due to the highly uniform intracavity intensity (as a result of the high mirror reflectivities) together with the fact that $\ln(1/R) \approx 1 - R$, so that the distributed out-coupling of the DLA is not a bad approximation of the true lumped nature of the mirror losses.

2. $\alpha_0 = 0$

When $\alpha_0 = 0$ there is no finite limiting value of the intracavity intensity $\beta^{\max}$; still, this is not enough of a reason to expect the DLA to be a good approximation because there can be significant intensity variation in the cavity–particularly for small R–that would seem to necessitate the nonuniform gain saturation treatment. Taking the limit of Eqs. 7 and 8 as α_0 goes to 0 yields

$$g_0 L = \ln\left(\frac{\beta_2^+}{\beta_1^+}\right) + \beta_2^+ - \beta_1^+ - \beta_0^2\left(\frac{1}{\beta_2^+} - \frac{1}{\beta_1^+}\right). \tag{A2}$$

For the FP cavity, this can be solved for the intensity emitted from each facet

$$\beta_1^{\text{out}} = (1 - R_1)\beta_1^- = \frac{(1 - R_1)L(g_0 - g_{0,th})}{1 - R_1 + \sqrt{R_1/R_2} - \sqrt{R_1 R_2}}, \tag{A3}$$

$$\beta_2^{\text{out}} = (1 - R_2)\beta_2^+ = \frac{(1 - R_2)L(g_0 - g_{0,th})}{1 - R_2 + \sqrt{R_2/R_1} - \sqrt{R_1 R_2}}, \tag{A4}$$

and the total emitted intensity

$$\beta^{\text{out}} = \beta_1^{\text{out}} + \beta_2^{\text{out}} = L(g_0 - g_{0,th}). \tag{A5}$$

For the ring cavity, it can be shown that the total emitted intensity is also given by Eq. A5 by setting $\beta_0 = 0$ in Eq. A2. Interestingly, the DLA predicts exactly the same output power (seen by setting $\alpha_0 = 0$ in Eq. 4) as the nonuniform gain saturation treatment in this limit. Again, this is not an obvious result, since the intracavity intensity variation can be made indefinitely large by reducing R. We note in passing that the two theories do not agree on the fraction of the total power coupled out of each mirror of the FP cavity, though the discrepancy is small in most cases. In particular, there is no discrepancy between the theories for the case of $R_1 = R_2$ (half of the power leaves from each mirror) or when one of the mirrors has unity reflectivity (all of the power is coupled out of the other mirror).

Appendix B: When is the asymptotic regime reached in the ring cavity?

In Sec. III A, we derived the output intensity of the ring cavity in Eq. 11. We took the limit of this equation for $g_0 \gg g_{0,th}$, resulting in Eq. 13, and called this the asymptotic regime because the output intensity approaches an asymptote. However, it turns out that g_0 does not even need to be too much greater than $g_{0,th}$ before the asymptotic regime is reached. To quantify how large g_0 must be, we look at the ratio of the output power given by the exact expression of Eq. 11 to that of the asymptotic expression of Eq. 13, $\beta^{\text{out}}/(\beta^{\text{out}}|_{g_0 \gg g_{0,th}})$, and define g_0^{asmp} as the gain above which this ratio is greater than 0.95. (Our choice of 0.95 is arbitrary, but is a good indicator for when the exact solution approaches the asymptotic solution.) In Figs. 7(a) and 7(b), $g_0^{\text{asmp}}/g_{0,th}$ is plotted against R for several different values of $\alpha_0 L$. The most striking conclusion is that provided $R > 0.01$ (a condition satisfied by nearly all practical lasers), the asymptotic theory becomes very accurate for $g_0 > 2.6 g_{0,th}$. In Fig. 7(a), we see that as $\alpha_0 L$ increases from 0.1 to 2 at fixed R, $g_0^{\text{asmp}}/g_{0,th}$ increases as well. However, for $\alpha_0 L \geq 3$ as shown in Fig. 7(b), this behavior changes, and $g_0^{\text{asmp}}/g_{0,th}$ instead decreases monotonically with increasing $\alpha_0 L$ at fixed R. Combining the information in Figs. 3 and 7(b), we reach an important conclusion. As $\alpha_0 L$ increases beyond ≈ 3, two things happen: 1) the asymptotic slope efficiency becomes significantly smaller than the threshold slope efficiency, and 2) the asymptotic regime is reached even closer to threshold.

Appendix C: Variation of intracavity intensity

The purpose of this appendix is to better understand the variation of the total intensity $\beta^+ + \beta^-$ within the cavity. For the ring cavity this is quite simple since $\beta^- = 0$, so the maximum and minimum occur at the facets, and their ratio is given by

$$\left.\frac{(\beta^+ + \beta^-)_{\min}}{(\beta^+ + \beta^-)_{\max}}\right|_{\text{ring}} = R. \tag{C1}$$

For the FP cavity, recall that $\beta^+\beta^- = \beta_0^2$, so the total intensity can be written as

$$\beta^+ + \beta^- = \beta^+ + \beta_0^2/\beta^+. \tag{C2}$$

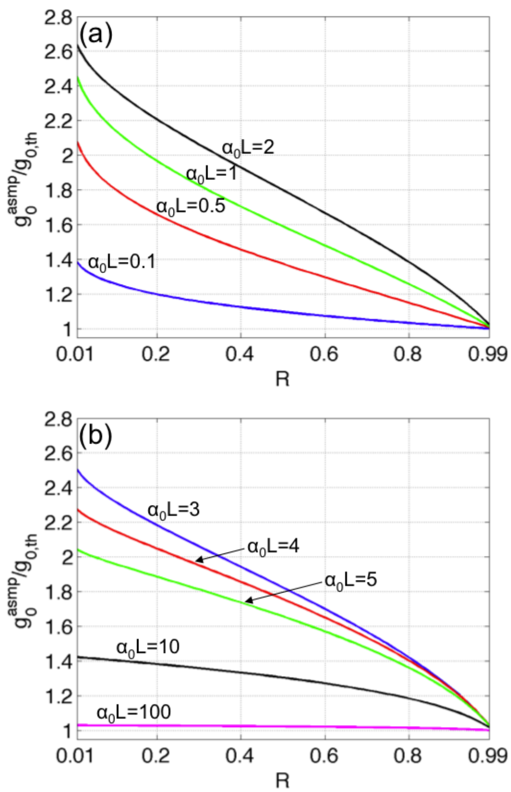

FIG. 7. The value $g_0^{\rm asmp}/g_{0,th}$–the gain at which the slope efficiency reaches 95% of its asymptotic value–is plotted for values of (a) $\alpha_0 L \leq 2$ and (b) $\alpha_0 L \geq 3$, which shows that $g_0^{\rm asmp}$ decreases monotonically with increasing $\alpha_0 L$ beyond $\alpha_0 L = 3$.

The minimum total intensity is found by differentiating Eq. C2 and occurs at the position where the two counter-propagating waves have equal intensity, i.e., $\beta^+ = \beta^- = \beta_0$, and therefore $(\beta^+ + \beta^-)_{\rm min} = 2\beta_0$. The maximum intensity can be found from Eq. C2 to occur at the facet of lower reflectivity, which we take to be R_2, and is given by $(\beta^+ + \beta^-)_{\rm max} = \beta_0(1 + R_2)/\sqrt{R_2}$. Thus, the total intensity variation in the FP cavity is given by

$$\left.\frac{(\beta^+ + \beta^-)_{\rm min}}{(\beta^+ + \beta^-)_{\rm max}}\right|_{\rm FP} = \frac{2\sqrt{R_2}}{1 + R_2}, \tag{C3}$$

where R_2 is the facet of lower reflectivity. It is interesting to note that this variation in total intensity is independent of g_0 or α_0–it is determined solely by the facet reflectivities. A ratio closer to 1 indicates a more uniform intensity in the cavity. Equation C3 is plotted in [11] (but was derived here under more general considerations). By comparing Eqs. C1 and C3, one finds that the intensity in the FP cavity is always more uniform than in the ring cavity, and in most cases substantially so. Physically, the reason for this is of course that the intensity in the FP cavity is shared between two counter-propagating waves which grow in opposite directions.

For fixed $R = \sqrt{R_1 R_2}$, the most uniform distribution is achieved when $R_1 = R_2$. As R_1 increases and R_2 decreases, the intensity becomes more nonuniform. The least uniform distribution occurs for $R_1 = 1$ and the minimum value of $R_2 = R^2$. This is the closest that the FP cavity intensity distribution can come to resembling the ring cavity intensity distribution–for a given R–because the right-propagating intensity β^+ is as large as possible relative to the left-propagating β^-. In this case, the ratio of Eq. C1 to Eq. C3 is $(1 + R_2)/2$, which tells us that for small R_2 the intensity of the FP cavity is still about twice as uniform as that of the ring cavity.

[1] A.E. Siegman, *Lasers*, 473-89 (1986).
[2] O. Svelto, *Principles of Lasers*, 5th Ed., Springer (2010).
[3] W.T. Silfvast, *Laser Fundamentals*, Cambridge University, Cambridge, 2004.
[4] A. Yariv and P. Yeh, *Photonics: Optical Electronics in Modern Communications*, 6th Ed., 246-48 (2005).
[5] B.E.A. Saleh and M.C. Teich, *Fundamentals of Photonics*, 2nd Ed. (2007).
[6] S.L. Chuang, *Physics of Photonic Devices*, 2nd Ed., 415-16 (2009).
[7] L.A. Coldren and S.W. Corzine, *Diode Lasers and Photonic Integrated Circuits*, 37-44 (1995).
[8] E.O. Schulz-DuBois, "Pulse sharpening and gain saturation in traveling-wave masers," Bell Syst. Tech. J. **43**, 625-658 (1964).
[9] W.W. Rigrod, "Gain Saturation and Output Power of Optical Masers," J. Appl. Phys. **34**, 2602 (1963).
[10] N.V. Karlov, *Lectures on Quantum Electronics*, 1st English Ed, 75-98 (1993).
[11] W.W. Rigrod, "Saturation Effects in High-Gain Lasers," J. Appl. Phys. **36**, 2487 (1965).
[12] W.W. Rigrod, "Homogeneously Broadened CW Lasers With Uniform Distributed Loss," IEEE J. Quant. Electron. **QE-14**, 377 (1978).
[13] G.M. Schindler, "Optimum Output Efficiency of Homogeneously Broadened Lasers With Constant Loss," IEEE J. Quant. Electron. **QE-16**, 546 (1980).
[14] T.R. Ferguson, "Lasers with saturable gain and distributed loss," Appl. Optics **26**, 2522 (1987).
[15] T.R. Ferguson and W.P Latham, "Efficiency and equivalence of homogeneously broadened lossy lasers," Appl. Optics **31**, 4114 (1992).
[16] L.W. Casperson, "Laser power calculations: sources of error." Appl. Optics **19**, 422 (1980).
[17] A.D. White, E.I. Gordon and J.D. Rigden, "Output power of the 6328-Å gas maser," Appl. Phys. Lett. **2**, 91 (1963).
[18] L.W. Casperson, "Threshold characteristics of multimode laser oscillators," J. Appl. Phys. **46**, 5194 (1975).
[19] L.W. Casperson and M. Khoshnevissan, "Threshold

characteristics of multimode semiconductor lasers," J. Appl. Phys. **75**, 737 (1994).

[20] C. Gmachl, F. Capasso, D.L. Sivco and A.Y. Cho, "Recent progress in quantum cascade lasers and applications," Rep. Prog. Phys. **64**, 1533 (2001).

[21] A. Wittmann, T. Gresch, E. Gini, L. Hvozdara, N. Hoyler, M. Giovannini and J. Faist, "High-Performance Bound-to-Continuum Quantum-Cascade Lasers for Broad-Gain Applications," IEEE J. Quant. Electron. **44**, 36 (2008).

Report of Referee 1

This is a good and useful paper, with interesting discussion and conclusions. My only reservation is that the paper largely ignores a rather long trail of previous publications on the subject specifically for the case of Fabry-Perot semiconductor lasers, in which longitudinal spatial hole burning is well known as one of the limitations to the output power. The first papers on the subject appeared very soon after semiconductor lasers themselves became an issue [Ulbrich1970]. Numerical and semi-analytical models have been developed, and verified experimentally, in papers such as [Fang1995, Rinner2002]. More recently, in [Ryvkin2011] it was shown fully analytically that in the absence of internal absorption, longitudinal spatial hole burning does not affect the output power of the laser at all – this ties up well with the conclusions in the present paper about the importance of internal absorption. Authors such as for example [Chen2012] have shown that indeed the effects of the spatial hole burning on the output power scale with the internal absorption. In an excellent review [6] the matter is also discussed in some detail. This is just a selection of papers, there are more. In the theoretical part of all the papers discussed, the models used are phenomenological so apply not just to semiconductor lasers but to all class B lasers. Some discussion of this cluster of literature is needed in the paper, and the claims of complete novelty need to be slightly tempered.

[1] ULBRICH, R; PILKUHN, MH, IEEE JOURNAL OF QUANTUM ELECTRONICS Volume: QE 6 Issue: 6 Pages: 314-& Published: 1970

[2] FANG, WCW; BETHEA, CG; CHEN, YK; et al. IEEE JOURNAL OF SELECTED TOPICS IN QUANTUM ELECTRONICS Volume: 1 Issue: 2 Pages: 117-128 Published: JUN 1995

[3]Rinner, F; Rogg, J; Friedmann, P; et a;., APPLIED PHYSICS LETTERS Volume: 80 Issue: 1 Pages: 19-21 Published: JAN 7 2002.

[4] Ryvkin, B. S.; Avrutin, E. A. JOURNAL OF APPLIED PHYSICS Volume: 109 Issue: 4 Article Number: 043101 Published: FEB 15 2011

[5] Chen, Zhigang; Bao, Ling; Bai, John; et al. Conference: Conference on Novel In-Plane Semiconductor Lasers XI Location: San Francisco, CA Date: JAN 23-26, 2012 NOVEL IN-PLANE SEMICONDUCTOR LASERS XI Book Series: Proceedings of SPIE Volume: 8277 Article Number: 82771J

[6] Wenzel, Hans IEEE JOURNAL OF SELECTED TOPICS IN QUANTUM ELECTRONICS Volume: 19 Issue: 5 Article Number: 1502913 Published: SEP-OCT 2013

Report of Referee 2:

In this manuscript, the authors discuss the slope efficiency of homogeneously broadened lasers with a non-uniform gain saturation as due to a spatially-dependent power distribution caused by both the internal and the lumped mirror losses. By analyzing the monochromatic solutions of the system, the authors show that, in the presence of internal losses, the output power vs pump characteristics is sublinear, in contrast with the predictions obtained in the Distributed Loss Approximation (DLA); as a result, the DLA may substantially overestimate the power output of the laser. The authors discuss the impact of cavity asymmetry, and they also briefly discuss the implications of their results on laser optimization and characterization, with particular emphasis on Quantum Cascade Lasers.

The manuscript is clearly written, the results are well presented and discussed, and the topic addressed is in my opinion well suited to Optica and of interest to the community involved in laser research.

However, the authors restrict themselves to the particular case where the self- and cross-saturation coefficients are the same for the forward and backward waves. This is only the case when the gain is not able to develop a grating at half the optical wavelength due to Spatial-Hole Burning (SHB) caused by the standing wave effect, e.g., when diffusion is large enough. Indeed, the authors explicitly state that they neglect SHB after eq. (1), but I think that this discussion would greatly benefit of further elaboration.

In fact, it has been recently shown [1] that, although small, the SHB grating has a profound impact on the modal stabilities, especially for nearly-symmetrical Fabry-Pérot cavities. In addition, it lifts the condition that $\beta_+(z)\beta_-(z)$=constant, which implies enormous complications for dealing with bidirectional cases. Clearly, for a fixed diffusion coefficient, the grating will be more important as the intensities grow, and at some point the validity of the starting equations will cease; if this occurs before the asymptotic regime is reached, the results presented here may not be observable. Some estimations of this “crossing point” would be most welcome.

I am not asking the authors to treat the complete problem, but I strongly recommend them to explicitly address these issues at the appropriate places in the manuscript, since in my opinion, it would greatly benefit the completeness and rigorousness of the work presented.

[1] Pérez-Serrano, A.; Javaloyes, J; Balle, S; «Longitudinal mode multistability in Ring and Fabry- Perot lasers: The effect of spatial hole burning»; Optics Express 19, p. 3284-3289 (2011)

Our response:

November 20th, 2014

Dear Professor Mork,

Thank you for the conditional acceptance of our manuscript. We thank the reviewers for their in-depth reading of our work and insightful comments. We have attached a highlighted version of our manuscript indicating all of the changes that we have made, which we will now discuss point by point.

Reviewer 1: My only reservation is that the paper largely ignores a rather long trail of previous publications on the subject specifically for the case of Fabry-Perot semiconductor lasers, in which longitudinal spatial hole burning is well known as one of the limitations to the output power. The first papers on the subject appeared very soon after semiconductor lasers themselves became an issue [Ulbrich1970]. Numerical and semi-analytical models have been developed, and verified experimentally, in papers such as [Fang1995, Rinner2002]. More recently, in [Ryvkin2011] it was shown fully analytically that in the absence of internal absorption, longitudinal spatial hole burning does not affect the output power of the laser at all – this ties up well with the conclusions in the present paper about the importance of internal absorption. Authors such as for example [Chen2012] have shown that indeed the effects of the spatial hole burning on the output power scale with the internal absorption. In an excellent review [Wenzel] the matter is also discussed in some detail. This is just a selection of papers, there are more. In the theoretical part of all the papers discussed, the models used are phenomenological so apply not just to semiconductor lasers but to all class B lasers. Some discussion of this cluster of literature is needed in the paper, and the claims of complete novelty need to be slightly tempered.

Response: We thank Reviewer 1 for directing us to this literature. Apparently in diode lasers "spatial hole burning" refers very generally to the nonuniformity of the intracavity intensity, while those of us in the QCL community use it to refer only to the intensity variation caused by standing-wave effects. Therefore, we were unaware of this body of work. While all of the work focuses on understanding the gain and intensity variation inside a cavity, only [Ryvkin2011] and [Chen2012] examine how this affects the power output of the laser. [Ryvkin2011] only deals with the case of zero distributed loss. [Chen2012] is a great paper, and even demonstrates the sublinear power output, but they provide only a very vague understanding of the result in terms of a "higher effective optical loss." We therefore emphasize in our manuscript that the novelty of our work is the simple explanation in terms of β^{max}, and also providing simple formulas to allow researchers to know when they must account for nonuniform gain saturation. We have included the following paragraph in our introduction:

Much work has been done on diode lasers–the prototypical example of a laser with low facet reflectivities–to understand the nonuniformity of the gain and intracavity intensity. In this context the nonuniformity is often referred to as (long-range) spatial hole burning (SHB), but we prefer to reserve the term "SHB" to describe only the (short-range) nonuniformity due to standing-wave effects. An early theoretical work to understand the nonuniformity was given in [16], and experiments have confirmed the expected longitudinal gain distribution [17, 18]. The fact that a nonuniform intensity distribution, in the absence of distributed loss, does not affect the power output was noted in [19], and we will generalize this result. The impact of distributed loss on the slope efficiency for a few specific cases was well-illustrated both computationally and experimentally in [20], but the understanding of the effect was limited. Our goal is to provide a physical understanding of the sublinear power output, which is a simple consequence of the intracavity intensity limit imposed by the distributed loss.

[Ryvkin2011] showed that the power output is not affected by nonuniform gain saturation if there is no distributed loss. However, he neglects nonradiative recombination of the electrons, which makes this result easier to understand. Our result allows for nonradiative recombination, which makes the result more surprising. We discuss this in the Supplemental Material.

We have also added a paragraph in the conclusion to discuss the assumption of uniform current density, and to draw attention to the suggestions by [Chen2012] to mitigate the sublinearity:

Our assumption of uniform small-signal gain g0 along the length of the laser is a common one. In electrically injected lasers, this assumption relies on a uniform current injection. This is not necessarily the case in highly-efficient lasers, if the electrical resistance of the laser decreases significantly with an increasing rate of stimulated emission. In such a laser, the current density would concentrate in the regions where the gain is most highly saturated. This effectively delivers electrons to where they are needed most, and would oppose the limiting influence of β^{max}*. One could also intentionally deliver more current to the region of highest intensity using multiple electrical contacts [20], or taper the waveguide width to dilute the optical intensity and prevent it from nearing* β^{max} *[20]. Both strategies would mitigate the sublinear power output.*

Reviewer 2: The authors restrict themselves to the particular case where the self- and cross-saturation coefficients are the same for the forward and backward waves. This is only the case when the gain is not able to develop a grating at half the optical wavelength due to Spatial-Hole Burning (SHB) caused by the standing wave effect, e. g., when diffusion is large enough. Indeed, the authors explicitly state that they neglect SHB after eq. (1), but I think that this discussion would greatly benefit of further ellaboration.

In fact, it has been recently shown [Perez-Serrano2011] that, although small, the SHB grating has a profound impact on the modal stabilities, especially for nearly-symmetrical Fabry-Pérot cavities. In addition, it lifts the condition that $\beta+(z)\beta-(z)$=constant, which implies enormous complications for dealing with bidirectional cases. Clearly, for a fixed difussion coefficient, the grating will be more important as the intensities grow, and at some point the validity of the starting equations will cease; if this occurs before the asymptotic regime is reached, the results presented here may not be observable. Some estimations of this «crossing point» would be most welcome.

I am not asking the authors to treat the complete problem, but I strongly recommend them to explicitly address these issues at the appropriate places in the manuscript, since in my opinion, it would greatly benefit the completeness and rigorousness of the work presented.

Response: We agree completely with Reviewer 2 that the mathematics will be complicated substantially by accounting for SHB and multimode operation. This is why we chose to stick with the simple case to demonstrate the sublinear output! However, while it would be difficult to ascertain the exact output power vs. pump power curve in such cases, we feel confident that the sublinearity will remain. This is because the limiting influence of β^{max} is very general: it should not matter how many complicated things are going on inside the cavity—a saturable gain medium with nonsaturable loss will impose a fundamental limit on the output power. Because this effect is not seen at threshold, the power output will first grow with the slope efficiency predicted by the uniform gain saturation theory, then decrease to comply with this fundamental limit. That is all that is needed to observe the sublinearity, and it in fact has been observed by [Chen2012]. We have added the following paragraph in our conclusion to address this point:

While our derivation explicitly treats only a single-mode laser and neglects SHB, in fact we expect the sublinear power output to be a quite general effect. To be completely rigorous, the standing-wave effects of SHB should be accounted for, and in a multimode laser the gain cross-saturation of each frequency on the others complicates the mathematics significantly. Therefore, the value of β^{max} *for each mode will be different and difficult to calculate. Nevertheless, the limiting effect of* β^{max} *will be small near threshold and substantial far above threshold, as we have shown. This is the essential ingredient for a sublinear power output.*

Furthermore, we have added a few curves to the plot of Fig. 4(a) to better explain the effect of the limiting intracavity intensity on the output power. We have also separated the appendices out into a supplementary file. Additionally, we wish to change the title of the manuscript to "Lasers with distributed loss have a sublinear output power characteristic" to emphasize the importance of distributed loss.

Thank you for considering our manuscript for publication in Optica.