A. Pick,1,* A. Cerjan,2 D. Liu,3 A. W. Rodriguez,4 A. D. Stone,2 Y. D. Chong,5 and S. G. Johnson6,†

1Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA
2Department of Applied Physics, Yale University, New Haven, Connecticut 06520, USA
3Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
4Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, USA
5Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore
6Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

DOI: 10.1103/PhysRevA.91.063806 PACS number(s): 42.55.Ah, 42.50.Lc, 42.60.Mi, 42.55.Sa

1. INTRODUCTION

The fundamental limit on the linewidth of a laser is a foundational question in laser theory [1–5]. It arises from quantum and thermal fluctuations [6,7] and depends on many parameters of the laser (materials, geometry, losses, pumping, etc.); it remains an open problem to obtain a fully general linewidth theory. In this paper, we present a multimode laser linewidth theory for arbitrary cavity structures and geometries that contains nearly all previously known effects and also finds new nonlinear and multimode corrections. The theory is quantitative and makes no significant approximations; it simplifies, in the appropriate limits, to the Schawlow-Townes formula (2) with the well-known corrections. It also demonstrates the interconnected behavior of these corrections [13,14], which are usually treated as independent. Most previous laser linewidth theories have employed simple models for calculating the lasing modes (e.g., making the paraxial approximation). Such simplifications, though appropriate for many macroscopic lasers, are inadequate for describing complex microcavity lasers such as three-dimensional (3D) nanophotonic structures or random lasers with inhomogeneities on the wavelength scale [15–18]. We base our theory on the recent steady-state ab initio laser theory (SALT) [19,20], which allows us to efficiently solve the semiclassical laser equations in the absence of noise for arbitrary structures [21]. We treat the noise as a small perturbation to the SALT solutions, allowing us to obtain the linewidths analytically in terms of simple integrals over the steady-state lasing modes. Our SALT-based theory is ab initio in the sense that it produces quantitatively accurate formulas for the linewidths, with no free parameters, including the full spatial degrees of freedom of the system. Hence, we refer to this approach as the noisy steady-state ab initio laser theory (N-SALT).

Our derivation (Secs. III–V) begins with the Maxwell-Bloch equations (details in Appendix A), which couple the full-vector Maxwell equations to an atomic gain medium [22], combined with random currents (in Sec. IV) whose statistics are described by the fluctuation-dissipation theorem (FDT) [23–27]. In the presence of these random currents, the amplitudes of the lasing modes evolve according to a set of coupled ordinary differential equations (ODEs), which have been called “oscillator models” [28,29] or “temporal coupled-mode theory” (TCMT) [30–34] in similar contexts. In their most general form, our N-SALT TCMT equations (Sec. III) have the form of oscillator equations with a noninstantaneous nonlinear term that stabilizes the mode amplitudes around their steady-state values. The noninstantaneous nonlinearity arises since the atomic populations respond with a time delay to field fluctuations; this corresponds to the typical case of “class B” lasers [35–37], in which the population dynamics cannot be adiabatically eliminated. We are able to show analytically that the resulting linewidths of the lasing peaks are identical to the results one obtains for a simplified model with instantaneous nonlinearity [28,29], which describes the (less common) case of “class A’ lasers, in which the population dynamics are adiabatically eliminated. As expected, however, in certain parameter regimes the full noninstantaneous model can exhibit side peaks alongside the main lasing peaks [38], arising from relaxation oscillations (Sec. V C).

By solving the N-SALT TCMT equations, we obtain a simple closed-form matrix expression for the linewidths and multimode phase correlations (Sec. V), generalizing earlier two-mode results that used phenomenological models [39]. This gives a multimode “Schawlow-Townes” relation (Sec. VIC), where the linewidth of each lasing mode is proportional to a sum of inverse output powers of the neighboring
lasing modes. The theory is valid well above threshold, and whenever a new mode turns on, this inverse-power relation produces a divergence due to the failure of the linearization approximation near threshold. However, we show that this divergence is spurious and can be avoided by solving the nonlinear N-SALT TCMT equations numerically [40]. (Our formalism can be extended to treat the near-threshold regime analytically by including noise from subthreshold modes, as discussed in Sec. VI B and in Sec. VIII.) Sections VI and VII also present several other model calculations that illustrate the differences between N-SALT and previous linewidth theories. Finally, in Sec. VIII, we discuss some potential additional corrections that will be addressed in future work. In a second paper [41], we also compare the theory against full time-dependent integration of the stochastic Maxwell-Bloch equations and find excellent quantitative agreement with the major results presented here.

Laser dynamics are reviewed in many sources [1–5], but it is useful to review here a simple physical picture of linewidth physics. A resonant cavity [e.g., light bouncing between two mirrors or a photonic-crystal (PhC) microcavity as in Fig. 1(a)] traps light for a long time in some volume, and lasing occurs when a gain medium is “pumped” to a population “inversion” of excited states to the point (threshold) where gain balances loss. [Of course, this simple picture is modified once additional modes reach threshold or for lasers (such as random lasers [42,43]) in which the passive cavity possesses no strong resonances; all of these complexities are handled by SALT [19,20] and hence are incorporated into our approach.] For simplicity, consider here a laser operating in the single-mode regime. Above threshold, the gain depends nonlinearly on the mode intensity |a|², as sketched in Fig. 1(b): Increasing the field intensity decreases the gain due to depletion of the excited states until it reaches a stable steady-state value a₀². (This gain-saturation effect is called “spatial hole burning” since it can be spatially inhomogeneous.) In the absence of noise, this results in a stable sinusoidal oscillation with an infinitesimal linewidth, but the presence of noise, which can be modeled by random current fluctuations J [10,29,44], perturbs the mode as depicted in Fig. 1(c), resulting in a finite linewidth. There are various sources of noise in real lasers, but spontaneous emission sets a fundamental lower limit on the linewidth [4]; here we include only spontaneous emission and thermal noise. In particular, although the squared amplitude is stabilized around a₀² by the nonlinear gain, the phase ϕ of the mode drifts according to a random walk (a Brownian or Wiener phase) with variance ⟨ϕ²⟩ ≈ Γ τ, and the Fourier transform of a Wiener phase yields a Lorentzian line shape [Fig. 1(d)] with full width at half maximum (FWHM) Γ [28]. The goal of linewidth theory is to derive Γ, ideally given only the thermodynamic FDT description of the current fluctuations and the Maxwell-Bloch physics of the laser cavity.

The most basic approximation for the linewidth (sufficiently far above threshold), usually referred to as the Schawlow-Townes (ST) formula [6,7], takes the form

$$\Gamma = \frac{\hbar \omega_0 \gamma_0}{2P}$$

where P is the output power of the laser, γ₀ is the passive-cavity resonance width, and ω₀ is the laser frequency, often approximated to be equal to the real part of the passive-cavity resonance pole at ω₂ = ω₀ + i γ₀/2. (A slightly more accurate approximation for the laser frequency takes into account the small line pulling of the laser frequency towards the atomic transition frequency [45].) The inverse-power dependence causes the famous line-narrowing of a laser above threshold.

Over the decades, a number of now-standard corrections to this formula were found [3–5], leading to the modified ST formula:

$$\Gamma = \frac{\hbar \omega_0 \gamma_0}{2P} \left[ \frac{n_p}{n_{sp}} \left( \frac{\gamma_\perp}{\gamma_\perp + \frac{\gamma_\parallel}{2}} \right)^2 \left( 1 + a_0^2 \right) \right]$$

First, the gain medium can be thought of, in many respects, as a system at negative temperature T [46], with the limit of complete inversion of the two lasing modes corresponding to T → 0⁻. When only partial inversion is present, the linewidth is enhanced by a factor of nₚ = n₀/γ₀ [47,48], where N₂ and N₁ are the spatially averaged populations in the upper and lower states of the lasing transition. We refer to this correction as the incomplete-inversion factor (also known as “the spontaneous emission factor”). Second, due to the openness of the laser system, the modes are not power orthogonal and the noise power which goes into each lasing mode is enhanced [49]; this correction is known as the Petermann factor, and it becomes significant in low-Q laser systems, where it is not a good approximation to treat the lasing mode E₀ as purely real. (Q ≡ ω₀/γ₀ is a dimensionless passive-cavity lifetime defined in units of the optical period [33].) Note that E₀ is the passive-cavity mode [in contrast to SALT solutions, which are the modes of the full nonlinear equations, introduced in (6)]. ∫₀𝐸₀ dx denotes integration over the cavity region. Third, for low-Q laser cavities, it is possible that the gain linewidth γₚ can be on the order of or smaller than the passive-cavity resonance width γ₀, causing significant...
dispersion effects as the gain is increased to threshold [9]. This correction is commonly called the “bad-cavity” factor [10,50]. Unlike the other corrections mentioned above, the bad-cavity factor decreases the laser linewidth. However, very few laser systems are in the parameter regime where this effect is significant [51]. Finally, amplitude fluctuations in the laser field couple to the phase dynamics, leading to a correction known as the “α factor.” For atomic gain media, this effect was identified by Lax [9] in the 1960s, and for this case it is typically a small correction. For bulk semiconductor gain medium the effect is large and typically dominates the broadening due to direct phase fluctuations [52–54]; in this context it is known as the “Henry α factor” [11].

Previous linewidth derivations have taken a number of different approaches, making severe approximations compared to the solution of the full 3D space-dependent Maxwell-Bloch equations in the presence of noise. Generally speaking, linewidth theories can be classified into two categories. The first class includes methods which solve Maxwell’s equations with a phenomenological model for the gain medium and account for noise spatial and spectral correlations by using the FDT [10,29,44]. Typically, these methods do not handle nonlinear spatial hole-burning above threshold or multimode effects. These methods, commonly used in the semiconductor laser literature, resulted in linewidth formulas which included the Petermann [49], bad-cavity [2,10], incomplete-inversion [29], and α factors [11]. Most notably, an early work by Arnaud [55] derived a single-mode linewidth formula without making any simplifying assumptions about the field patterns, handling anisotropic, inhomogeneous, and dispersive media. However, this theory was only applied to very simple, effectively 1D, homogeneous systems, and it was missing hole-burning effects and the α factor.

The second class of linewidth theories consists of scattering-matrix methods [13,14,56,57], which can treat arbitrary geometries without phenomenological parameters and take into account the effects of spatial hole burning. S-matrix theories only have access to the input and output fields and, therefore, can only treat the noise in a spatially averaged manner and are not able to obtain the α factor rigorously. However, they obtain all of the other corrections to the single-mode linewidth. In particular, the recent S-matrix approach by Chong et al. [13,14] takes advantage, as we do, of the ab initio computational approach of SALT and hence has the potential to treat arbitrary geometries and spatial hole-burning effects. (We reduce our results to the most recent scattering-matrix linewidth formula [14] in Appendix D.) Note that, in practice, S-matrix methods require a substantial independent calculation beyond SALT to extract the linewidths, whereas our approach obtains the linewidths immediately from SALT calculations (or any other method to obtain the steady-state lasing modes) by simple integrals over the fields.

Our derivation of N-SALT, being based on the SALT solutions, has a similar regime of validity. For single-mode lasing, SALT and N-SALT are essentially exact, relying only on the rotating-wave approximation and on the laser being sufficiently far above threshold. For multimode lasing, those theories require two additional dynamical constraints [19,20]: The rates associated with population dynamics must be small compared to both the dephasing rate of the polarization and the lasing-mode spacing (roughly, the free spectral range). The former constraint is satisfied in all solid-state lasers, whereas the latter requires a sufficiently small laser cavity. The actual size depends on details of both the cavity and the gain medium used, but the appropriate limit is realized in many complex lasers of interest. When these frequency scales are not well-separated, the level populations are not quasistationary, and multimode SALT will initially lose accuracy and eventually fail completely (since multimode lasing becomes unstable [58]). Moreover, while the average (SALT) behavior is unaffected by nonlasing poles, they do affect the noise properties, and N-SALT in its current form only accounts for a finite number of poles in the Green’s function (Appendix A.2). [We only include lasing poles (i.e., poles on the real axis), but extension to include nonlasing poles, which determine the amplified spontaneous emission (ASE) [40,59], is straightforward (Sec. VIII).] As noted above, the linewidth formula additionally assumes that the laser is operating far enough above threshold that amplitude fluctuations are small compared to the steady-state amplitudes (i.e., |a(t)| ≈ a0 in the notation of Sec. V). Hence, our formula does not describe the linewidth near the lasing thresholds. Our perturbation approach takes into account only the lowest-order correction to the complex modal amplitude a(t) and neglects higher-order corrections to the frequency ω0 and spatial pattern E0(x) [see Eq. (7)]. Moreover, we neglect non-Lorentzian corrections to the line shape [60–64] (Sec. IV). In the following section we present our generalized linewidth formula in the single-mode regime (3) and compare it with traditional linewidth theories.

II. THE N-SALT linewidth formula

Our main result is a multimode linewidth formula which generalizes (2). In the multimode case, the result takes the form of a covariance matrix for the phases of the various modes, which is presented in (36) and (37) of Sec. V. In the single-mode case, the N-SALT linewidth formula takes the simple form

$$\Gamma = \frac{\hbar \omega_0^2}{2p} \tilde{\epsilon}_{sp} K \tilde{L}(1 + \tilde{\alpha}^2).$$

(3)

The modified correction factors (marked by tildes) are defined in Table I. As can be seen from the table, those factors generalize the traditional expressions by taking into account both spatial inhomogeneity and nonlinearity. Since the generalized factors depend on the SALT permittivity ε, mode profile E0(x), and frequency ω0, one can no longer regard the effects of cavity openness, nonlinearity, and dispersion as separate multiplicative effects. In this sense, our formula demonstrates the intermingled nature of the linewidth correction factors, as previously introduced in [13,14], but here demonstrated at a new level of generality. We denote by ∫ dx integration over all space, for any number of spatial dimensions. We use the shorthand notation for vector products |E0|^2 = E0 ⋅ E0* and E0* = E0 ⋅ E0, where the latter unconjugated inner product appears naturally because of the biorthogonality relation for lossy complex-symmetric systems [65,66]. Im(ε(x)) denotes the imaginary part of the nonlinear steady-state permittivity (5), which is negative (positive) in gain (loss) regions.
TABLE I. Traditional and new linewidth correction factors for the single-mode linewidth formulas (2,3).

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Traditional</th>
<th>Generalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_0$</td>
<td>$\gamma_0$</td>
<td>$\int dx \left( \omega_0 \Im \varepsilon \right)</td>
</tr>
<tr>
<td>cavity decay rate</td>
<td>$\gamma_0$</td>
<td>$\int dx \left( \omega_0 \Im \varepsilon \right)</td>
</tr>
<tr>
<td>$\tilde{n}_{np}$</td>
<td>$N_2 - N_1$</td>
<td>$\int_p dx \Im \varepsilon</td>
</tr>
<tr>
<td>incomplete inversion</td>
<td>$\frac{N_2}{N_2 - N_1}$</td>
<td>$\left( \frac{N_2}{N_2 - N_1} \right)^2$</td>
</tr>
<tr>
<td>$\tilde{K}$</td>
<td>$\frac{\int_c dx</td>
<td>E_0</td>
</tr>
<tr>
<td>bad cavity</td>
<td>$\left( \gamma_\perp + \frac{\omega}{2} \right)^2$</td>
<td>$\frac{\int dy \varepsilon</td>
</tr>
<tr>
<td>$\tilde{\alpha}$</td>
<td>$\frac{\omega_0 - \omega_0}{\gamma_\perp}$</td>
<td>$\frac{\Im C}{\Re C}$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$\frac{1}{\hbar \omega_0} \ln \left[ \frac{N_1(x)}{N_2(x)} \right]$</td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td>$\frac{-i \frac{\omega}{2} \int dx \frac{\omega \varepsilon}{</td>
<td>E_0</td>
</tr>
</tbody>
</table>

The output power $P$ is related to the SALT solutions by invoking Poynting’s theorem, which can also be shown to use that $P \propto \int_0^{\infty} dx \left[ -\Im \varepsilon(x) \right] |E_0(x)|^2$. We use $\int_0^{\infty} dx$ to denote some volume which contains the gain medium. The choice of the volume is somewhat arbitrary; e.g., integrating over the cavity region corresponds to the output power at the cavity boundary [29]. Note, however, that this arbitrariness in the choice of the volume is not a general feature of our formula. After substituting the relevant expressions from Table I into (3), the integrals which contain $\int_0^{\infty} dx$ cancel, resulting in an expression for the linewidth only in terms of integrals over the entire space. The effective inverse temperature $\beta(x)$ is determined by the inhomogeneous steady-state atomic populations $N_1(x)$ and $N_2(x)$ and is defined as [67–69]

$$\beta(x) \equiv \frac{1}{\hbar \omega_0} \ln \left[ \frac{N_1(x)}{N_2(x)} \right].$$

In regions where the gain medium is pumped sufficiently to invert the population, $\beta(x)$ is negative; in regions where the pump is too weak to invert, $\beta(x)$ will be positive [and still given by (4)]; and in unpumped regions, Eq. (4) will simply reduce to the equilibrium temperature $(k_B T)^{-1}$ of the surrounding environment. The quantities $N_1(x)$ and $N_2(x)$ are an output of the SALT solution in the absence of noise. The spatially dependent expression inside the square brackets in the definition of $\tilde{n}_{np}$ in Table I generalizes the spatially averaged incomplete-inversion factor $\frac{N_2}{N_2 - N_1}$. That can be seen by noting that

$$\frac{1}{2} \coth \left( \frac{\omega_0}{2} \right) \frac{1}{2} = (\exp[\hbar \omega/\beta] - 1)^{-1} \approx n_B,$$

where $n_B$ is the usual Bose-Einstein distribution function [70,71]. (For gain media, it is sometimes convenient to introduce the positive spontaneous-emission factor $n_{sp} = -n_B$ [72]. Note that this definition ensures that the generalized incomplete-inversion factor is always positive.) The $\frac{1}{2}$ factor subtracted from the hyperbolic cotangent was discussed in [72], and we give a simple classical explanation for it in Appendix E. If standard absorbing layers are used to implement outgoing boundary conditions in the SALT solver [21] and the temperature of the ambient medium is assigned to these layers, then the N-SALT formula includes the effect of incoming thermal radiation. A generalized Petermann factor which formally resembles $K$ appeared in previous work by Schomerus [57] (in his expression for the Petermann factor for TM modes in 2D dielectric resonators). However, the earlier formula was expressed in terms of passive resonance scalar fields, whereas our correction contains 3D nonlinear SALT solutions. Finally, $\tilde{\alpha}$ is a generalized $\alpha$ factor, defined explicitly in Sec. V (30).

For atomic gain media, the traditional factor is expressed in terms of the atomic transition frequency $\omega_0$ and decay rate of the atomic polarization $\gamma_\perp$. In the current work we only evaluate the atomic case, although the general expression in terms of the nonlinear coupling $C$ should also apply to the semiconductor case.

The N-SALT formula (3) reduces to the traditional formula (2) in some limiting cases. Let us consider, for simplicity, a 1D Fabry-Pérot laser cavity of length $L$ surrounded by air (i.e., $Im \varepsilon = 0$ outside the cavity region). Let us assume also that the laser is operating not too far above the threshold and is uniformly pumped; hence, $Im \varepsilon$ and $\beta$ are nearly constant inside the cavity. In this limit, all the integrals in Table I can be approximated by reducing the integration limits to the cavity region; terms which contain integration over the imaginary part of the permittivity are nonzero only within the cavity region (e.g., $\int dx \Im \varepsilon |E_0|^2$); while terms of the form $\int dx \varepsilon |E_0|^2$ can be written as the sum of the cavity contribution $\int_c dx |E_0|^2$ and the surrounding medium contribution $\int_{\lim}^{\infty} dx |E_0|^2$, where the latter is negligible for $L \omega_0 \gg 1$, as shown in Appendix D and in [14] (here and throughout the paper, we set $c = 1$). Using this approximation, it is immediately apparent from Table I that the incomplete-inversion factor reduces to the traditional expression. The generalized Petermann factor reduces to the traditional factor in the limit of a high-$Q$ cavity, where the threshold lasing state $E(x)$ is approximately the same as the passive resonance state $E(x)$. In order to simplify the remaining terms, recall that the lasing threshold is reached when gain in the system compensates for the loss. For weak losses (small Im $\varepsilon/\beta$) that can be treated by perturbation theory, the threshold condition is $\gamma_0 = \omega m \varepsilon/\beta$ [2] and, therefore, the generalized decay rate reduces to $\gamma_0$ (one can thereby see that the ST formula (2) neglects nonlinear corrections to $\gamma_0$ as was also shown in [13]). Next, let us discuss the generalized bad-cavity factor, which simplifies to $(1 + \frac{\omega_0}{2 \gamma_0})^{-1}$ after reducing the integration limits. In order to show that it agrees with the traditional factor, we need to show that $\frac{\omega_0}{2 \gamma_0} \approx \frac{\omega_0}{2 \gamma_0}$. The steady-state effective permittivity, as used in SALT theory (Appendix A 1), is

$$\varepsilon(x) = \varepsilon_c(x) + \frac{\gamma_\perp D(x)}{\omega_0 - \omega_0 + i \gamma_\perp},$$

where $\varepsilon_c$ is the passive permittivity and the second term is the active nonlinear permittivity due to the gain medium. The population inversion $D(x) = N_2(x) - N_1(x)$ is generally spatially varying above threshold due to spatial hole burning.
Since we assume here that we are close to threshold and that the pumping is uniform, the inversion is also uniform in space and near its threshold value. If one assumes, additionally, that the detuning of the lasing frequency from atomic resonance is small (|ω0 − ωa| ≪ γ⊥), one obtains \[ \frac{\partial \alpha}{\partial \omega_{\text{res}}} \approx \text{Im} \varepsilon \frac{\partial \omega}{\partial \omega_{\text{res}}} \]. Finally, we show in Sec. VI.A that our \( \tilde{\alpha} \) reduces to the known \( \alpha_0 \) in homogeneous low-loss cavities, so that all factors of the corrected ST formula are recovered in this limit. (Note that line-pulling effects which may modify the lasing frequency \( \omega_0 \) are handled by SALT.)

In the next section, we present the TCMT equations which are used in this paper to derive the N-SALT linewidth formula (3), but which may also be used to extract more information on laser dynamics away from steady state.

### III. THE N-SALT TCMT EQUATIONS

In the absence of noise, the electric field of a laser operating in the multimode regime is given by the real part of \( \mathbf{E}_0(x,t) \), where

\[
\mathbf{E}_0(x,t) = \sum_{\mu} \mathbf{E}_\mu(x) a_{\mu0} e^{-i \omega_{\mu0} t},
\]

and the laser has zero linewidth. (This assumes, of course, that there exists a steady-state multimode solution of the nonlinear semiclassical lasing equations [19,20].) The modes \( \mathbf{E}_\mu(x) \) and frequencies \( \omega_{\mu0} \) can be calculated using SALT, which solves the semiclassical Maxwell-Bloch equations in the absence of noise. (SALT has been generalized to include multilevel atoms [73], multiple lasing transitions, and gain diffusion [74]; any of these cases can thus be treated by N-SALT with minor modifications, but we focus on the two-level case here.) The linewidth can now be calculated by adding Langevin noise, as described below.

In the presence of a weak noise source, the electric field can be written as a superposition of the steady-state lasing modes with time-dependent amplitudes \( a_{\mu}(t) \) which fluctuate around \( a_{\mu0} \):

\[
\mathbf{E}(x,t) = \sum_{\mu} \mathbf{E}_\mu(x) a_{\mu}(t) e^{-i \omega_{\mu0} t}.
\]

In principle, the sum in (7) should also include the nonlasing modes since the set of lasing modes by itself does not form a complete basis for the fields. Nonlasing modes contribute to ASE, which has a significant effect on the spectrum near and below the lasing thresholds [40,59] and will be treated in future work.

In Appendix A, we derive the N-SALT TCMT equations of motion for \( a_{\mu}(t) \) starting with the full vectorial Maxwell-Bloch equations. We show that the noise-driven field obeys an effective nonlinear equation which, in the frequency domain, takes the form

\[
\left[ \nabla \times \nabla \times - \omega^2 \varepsilon(\omega,a) \right] \tilde{\mathbf{E}}(x,\omega) = \tilde{\mathbf{F}}_s(x,\omega),
\]

where the carets denote Fourier transforms [e.g., \( \mathbf{E}(x,t) \equiv \int_{-\infty}^{\infty} d\omega e^{-i\omega t} \tilde{\mathbf{E}}(x,\omega) \)]. Spontaneous emission is included via the stochastic noise term \( \tilde{\mathbf{F}}_s(x,\omega) \) (quantified in Sec. IV), and the effective permittivity \( \varepsilon(\omega,a) \) (derived in Appendix A2) is given by

\[
\varepsilon(\omega,a) \tilde{\mathbf{E}}(x,\omega) = \sum_{\mu} \left[ \varepsilon a_{\mu} + \frac{\gamma^\perp}{\omega - \omega_a + i \gamma^\perp} \tilde{D} \right] \mathbf{E}_\mu(x),
\]

where the asterisk denotes a convolution. The second argument of \( \varepsilon(\omega,a) \) denotes the implicit dependence of \( \varepsilon \) on the modal amplitudes \( a_{\mu} \) through the inversion \( \tilde{D} \). The effective permittivity (9) can be decomposed into a steady-state-amplitude dispersive term and a nonlinear nondispersive term (similar in spirit to [75]). The key point here is that, to lowest order, there are two corrections to the permittivity in the presence of noise: the dispersive correction due to any shift in frequency at the unperturbed amplitudes \( a_{\mu0} \) and the nonlinear correction due to any shift in amplitude at the unperturbed frequency. (Shifts in frequency are small because only frequency components within the mode linewidths matter, while shifts in amplitude are small because of the stabilizing effect of gain feedback.) The coupling between these two perturbations is higher order and is hence dropped, which greatly simplifies the analysis.

Substituting the permittivity expansion (derived explicitly in Appendix A3) into Maxwell’s equation (8), we find that the noise-driven field obeys the linearized equation

\[
[\nabla \times \nabla \times - \omega^2 \varepsilon(\omega,a)] \tilde{\mathbf{E}}(x,\omega) = \tilde{\mathbf{F}}_{NL}(x,\omega) + \mathbf{f}_s(x,\omega);
\]

i.e., the dispersive permittivity which appears on the left-hand side of (10) is evaluated at the steady-state amplitude \( a_{\mu0} \). The nonlinear nondispersive term \( \mathbf{f}_{NL} \) [defined explicitly in (A23)], which corresponds to amplitude fluctuations at the unperturbed frequency, appears as a restoring force on the right-hand side. The noise-driven field \( \tilde{\mathbf{E}}(x,\omega) \) is found in Appendix A4 by convolving the linearized Green’s function with the source terms \( \mathbf{F}_{NL} \) and \( \mathbf{f}_s \). Finally, the N-SALT TCMT equations are obtained by transforming the noise-driven field back into the time domain.

### A. Time-delayed multimode model

We find that, in the most general case, the TCMT equations take the form

\[
a_{\mu} = \sum_{\nu} \int d\mathbf{x} c_{\mu\nu}(\mathbf{x}) \times \left[ \gamma(\mathbf{x}) \int d\mathbf{t}' e^{-i\omega t (\mathbf{x},\mathbf{t}')} |a_{\nu}(\mathbf{t}')|^2 \right] a_{\mu} + f_{\mu}.
\]

Comparing (11) and (10), one can see that the first term on the right-hand side of (11) is related to the nonlinear restoring force \( \mathbf{F}_{NL} \) and that the Langevin noise \( f_{\mu}(t) \) is associated with \( \mathbf{f}_s \).

The nonlinear coupling coefficients \( c_{\mu\nu}(\mathbf{x}) \) [derived in (A33)] correspond to local changes in the nonlinear permittivity with respect to intensity changes in each of the modes

\[
c_{\mu\nu} = -\frac{i \alpha_{\mu0}^2 (\partial \alpha_{\nu0} / \partial \omega_a)}{\int d\mathbf{x} (\omega_a^2 |a_{\nu}(\mathbf{t})|^2)} |\tilde{\mathbf{E}}_{\mu0}(\mathbf{x})|^2.
\]
where we have introduced a shorthand notation for the derivative in the denominator \( \frac{\partial}{\partial \omega} \epsilon \bigg|_{\omega_0} \). This modulating coupling in the fluctuation dynamics comes about because of the saturation of the gain. A fluctuation in mode \( \mu \) affects the amplitudes of all the other modes \( \nu \).

The N-SALT TCMT equations are nonlocal in time because the atomic populations are not, in general, able to follow the field fluctuations instantaneously and, instead, respond with a time delay determined by the local atomic decay rate \( \gamma(x) \), given by

\[
\gamma(x) = \gamma \left[ 1 + \sum_{\nu} \frac{y^2_{\nu}}{(\alpha_\nu - \omega_\nu)^2 + \gamma^2} |a_\nu|^2 |E_x|^2 \right].
\] (13)

The second term in (13) is precisely the local enhancement of the atomic decay rate due to stimulated emission in the presence of the lasing fields. (A simplified spatially averaged enhancement of the atomic decay rate was previously discussed in [38].)

The Langevin force \( f_\mu \) is the projection of the spontaneously emitted field onto the corresponding mode \( E_\mu \). Defining \( F_\mu(t) \equiv F_S e^{i \omega_\mu t} \), the Langevin force \( f_\mu \) is

\[
f_\mu(t) = \frac{i}{\hbar} \int dx E_{\mu} \cdot F_\mu(t) \frac{dx}{(\omega^2_\mu - \omega^2)(\gamma^2)} E_\mu^2.
\] (14)

The full N-SALT TCMT equations (11) describe the most typical situation in laser dynamics of a “class B” laser [35–37], in which the polarization of the gain medium can be adiabatically eliminated but the population dynamics is relatively slow and cannot be so eliminated. However, much of the basic linewidth physics can be extracted from the limit when the population dynamics is also adiabatically eliminable, which describes “class A” lasers. Since the mathematical analysis is simpler in this limit, we begin the spectral analysis in Sec. V with the latter model. We discuss this limit, which we refer to as the “instantaneous model,” in the following section.

B. Instantaneous single-mode model

When the population relaxation rate \( \gamma(x) \) is (everywhere) large compared to the dynamical scales determining \( a_\mu(t) \), the exponential terms in (11) act like \( \delta \) functions. After the spatial integration, and specializing in this section to the single-mode case, we obtain the simple nonlinear oscillator model driven by a weak Langevin force \( f(t) \).

\[
\dot{a} = C(a^2_0 - |a|^2) a + f,
\] (15)

where \( C = \int dx c(x) \) is the integrated nonlinear coupling. This instantaneous nonlinear oscillator model was previously introduced by Lax [9,28] and has been used extensively in linewidth theories [2]. The N-SALT approach enables computing the model’s parameters \( ab initio \), taking full account of the spatial hole-burning term and the vectorial nature of the fields [including multimode effects, when generalizing (15) to the multimode regime]. Also, our approach shows that this well-known model can be explicitly derived from the more general (noninstantaneous) model, presented in the previous section. Above the lasing threshold, \( a_0 > 0 \) and \( \text{Re}[C] > 0 \), and the system undergoes self-sustained oscillations with a stable state at \(|a| = a_0 \), as demonstrated in Fig. 1(b). In fact, near threshold one can show that \( \text{Re}[C] \) is approximately the threshold gain, which balances the cavity loss \( \kappa \). Hence, the dynamical scale of \( a(t) \) is of order \( \kappa \), which must then be much smaller than \( \gamma(x) \) for the instantaneous model to hold: this is the standard dynamical condition for class A lasers [35–37].

The nonlinear term in (15) and the multimode counterpart in (12) are derived rigorously in Appendix A, but we can motivate the resulting expressions using simple physical arguments. The nonlinear term can be viewed as a shift in the oscillation frequency, i.e., \(-i \Delta \omega = C(a^2_0 - |a|^2) \). Using first-order perturbation theory [76], the frequency shift due to a change in dielectric permittivity \( \Delta \epsilon \) is given by

\[
\Delta \omega = -\omega^2_0 \int dx \Delta \epsilon |E^2_\mu| \frac{dx}{(\omega^2_\mu - \omega^2)(\gamma^2)} E^2_\mu.
\] (16)

Plugging in the differential of the permittivity due to small changes in the squared mode amplitude, \( \Delta \epsilon \approx \frac{\partial}{\partial |a|^2} (|a|^2 - a^2_0) \), we find that the coupling coefficient in the instantaneous model is

\[
C = -i \omega^2_0 \int dx \frac{\partial}{\partial |a|^2} (|a|^2 - a^2_0) |E^2_\mu| \frac{dx}{(\omega^2_\mu - \omega^2)(\gamma^2)} E^2_\mu.
\] (17)

This is the single-mode version of (12) integrated over space due to rapid relaxation. As we will see, this simple result, combined with the spectrum of the Langevin noise (Sec. IV), is all that is needed to derive the single-mode linewidth formula (3) (see Sec. V), and the multimode generalization also follows straightforwardly. Hence, after analyzing the noise spectrum, we first derive the linewidth within the instantaneous model before moving on to the more complicated case of the full N-SALT TCMT equations. The latter will show that the basic linewidth formula is unchanged from that of the instantaneous model except for the addition of side peaks due to the relaxation oscillations present in class B lasers.

IV. THE AUTOCORRELATION FUNCTION OF THE LANGEVIN FORCE

In this section, we express the autocorrelation function of the Langevin force \( f_\mu(t) \)

\[
(f_\mu(t)f^*_\nu(t')) = R_\mu R_\nu \delta_{\mu\nu} \delta(t - t')
\] (18)

in terms of the autocorrelation function of the noise source \( F_\mu \). It is well known that quantum and thermal fluctuations can be modeled as zero-mean random variables, defined by their correlation functions [26,27]. This Rylov picture [24] is essentially a consequence of the central-limit theorem (CLT) [77,78], which holds since the classical forcing \( F_S \) is the sum of a large number of randomly emitted photons. The autocorrelation function of \( F_S \) can be found by invoking the FDT, as explained below.

The probability distributions of the pumped medium and the electromagnetic field obey Boltzmann statistics, with an effective local temperature \( \beta \) defined in terms of the atomic inversion [46] (see definition in Sec. II). Under the typical conditions of local thermal equilibrium [23–27], dissipation by optical absorption must be balanced by spontaneous emission from current fluctuations \( J(x,t) \). One can then apply the FDT for the Fourier-transformed forcing
\[ \hat{F}_S(x, \omega) = -4\pi i \omega \tilde{J}(x, \omega) \] [79]:

\[ \langle \hat{F}_S(x, \omega) \rangle \hat{F}_S^*(x', \omega') \] 

\[ = 2\hbar \omega^4 \Im \varepsilon(x, \omega) \coth \left( \frac{\hbar \omega \beta(x)}{2} \right) \delta(x - x') \delta(\omega - \omega'). \] (19)

Using this result, we calculate the autocorrelation of the Langevin force \( \tilde{f}_\mu \) [i.e., the Fourier transform of (14), defined as \( \tilde{f}_\mu(\omega) \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \, e^{i\omega t} \, f_\mu(t) \)] and we obtain

\[ \langle \tilde{f}_\mu(\omega) \rangle \tilde{f}_\mu^*(\omega') = \tilde{R}_\mu(\omega) \delta(\omega - \omega') \delta_{\mu\nu}, \] (20)

where the frequency-domain autocorrelation coefficient is

\[ \tilde{R}_\mu(\omega) = 4\hbar \omega^4 \left[ \int dx \left| E_\mu(x) \right|^2 \Im \varepsilon(\omega) \left( \frac{\hbar \omega \beta}{2} - \frac{1}{2} \right) \right] \left( \int dx E_\mu^2(x) \right)^2. \] (21)

The \( \frac{1}{2} \) factor subtracted from the hyperbolic cotangent is explained in Appendix E and in [72].

The time-domain diffusion coefficient \( R_\mu \) can be found directly from (21) taking the inverse Fourier transform. For the common case of a small linewidth, \( \Im \varepsilon(\omega) \) and \( \coth(\frac{\hbar \omega \beta}{2}) \) are nearly constant for frequencies within the linewidth. This means, essentially, that the Langevin force \( f_\mu(t) \) can be treated as white noise. Consequently, one can approximate the diffusion coefficient in (21) by its value at \( \omega_0 \).

With this simplification, the time-domain diffusion coefficient in (18) is conveniently given by \( R_\mu = 2\pi \tilde{R}_\mu(\omega_0) \) [29].

More generally, however, including this frequency dependence corresponds to temporally correlated fluctuations, leading to non-Lorentzian corrections to the laser line shape [60–64]. These “memory effects” can be addressed using our approach (as discussed in Sec. VIII) and we plan to include them in future work.

V. THE LASER SPECTRUM

In this section, we calculate the laser spectrum using the N-SALT TCMT equations (11) and (15) and the noise autocorrelation function (20) and (21). We begin by showing that the phase of the lasing mode undergoes simple Brownian motion; consequently, the laser spectrum is a Lorentzian, with a width given by the phase-diffusion coefficient. In Sec. VA, we calculate the phase-diffusion coefficient (hence the linewidth) for the instantaneous model (15) and in Sec. VB, we outline the analysis for the time-delayed model (11), leaving the details of the derivation to Appendix B. More accurately, the spectrum of the time-delayed model consists of a central Lorentzian peak at the lasing resonance frequency and additional side peaks due to relaxation oscillations, which are present in class B lasers. The latter side peaks are the subject of Sec. VC.

A. Instantaneous single-mode model

The complex mode amplitude \( a(t) \) can be written in polar form as

\[ a(t) = [a_0 + \delta(t)] e^{i\phi(t)}. \] (22)

\( a_0 \) is the steady-state amplitude, while \( \delta \) and \( \phi \) are real amplitude and phase fluctuations. Substituting the modal expansion (22) in (15), defining

\[
A \equiv 2a_0^2 \Re C, \\
B \equiv 2a_0^2 \Im C, \quad \text{and} \quad \phi_0 = -\phi^2 + 2A, \\
\delta = -A\delta + f_R, \\
a_0\phi = -B\delta + f_1,
\]

where \( f_R \equiv \Re \{f\} \) and \( f_1 \equiv \Im \{f\} \). We check the approximation of \( |\delta| \ll |a_0| \) a posteriori and we find that it generally holds (as was also shown in [40]), except near threshold \( (a_0 \to 0) \), which is a case we discuss in Sec. VIC.

When the nonlinear coupling coefficient is real \( (B = 0) \), it is evident from (25) that the phase undergoes simple Brownian motion (i.e., it is a Wiener process) and hence the phase variance increases linearly in time. An oscillator with Brownian phase noise has a Lorentzian spectrum [80], and one can reproduce that result briefly as follows. The laser spectrum \( S_\omega(\omega) \) is given by the Fourier transform of the autocorrelation function of \( a(t) \): \n
\[ \langle a(t) a^*(0) \rangle \approx a_0^2 e^{-\frac{1}{2}\delta^2} e^{-(\phi(0) - \phi(t))}, \]

(26)

For a Wiener phase, whose variance is \( (\phi(\tau) - \phi(0))^2 = \Gamma |\tau| \), the Fourier transform of (26) is a Lorentzian whose central-peak width is \( \Gamma \) [28]. In passing from the first to the second step in (26), one neglects direct amplitude-fluctuation contributions (which are decoupled from the phase) as these only introduce broad-spectrum background noise, but do not affect the linewidth of the laser peak (we return to this point in Sec. VC). In passing from the second to the third step, one assumes that the phase is a Gaussian normal variable, which is justified as a consequence of the CLT.

It is well known that also in the general case of \( B \neq 0 \), the phase is a Wiener process, with a modified diffusion coefficient [11]. In order to calculate the phase variance explicitly, we solve (24) and (25) and obtain

\[ \delta(t) = \int_0^t e^{-\Lambda(t-\tau)} f_R(\tau) d\tau, \]

\[ a_0\phi(t) = -B \int_0^t \delta(\tau) d\tau + \int_0^t f_1(\tau) d\tau. \]

Substituting (27) into (28), using the autocorrelation function of \( f \) (21), and performing the integration, one obtains that the phase variance in the long-time limit is \( (\phi(t) - \phi(0))^2 = \frac{R}{2a_0^2} (1 + \frac{B}{A}) |t| \) (where terms growing more slowly than \( |t| \) were neglected, as explained in greater detail in Appendix B). Therefore, the linewidth is

\[
\Gamma = \frac{R}{2a_0^2} (1 + \tilde{\alpha}^2),
\]

(29)

where we have defined the generalized \( \alpha \) factor,

\[
\tilde{\alpha} = \frac{B}{A} = \frac{\Im C}{\Re C},
\]

(30)
with the nonlinear coefficient $C$ defined in (17). Substituting the autocorrelation function (21) in (29) and using Poynting’s theorem to relate $a_0^2$ to the output power $P = \frac{\alpha a_0^2}{2\pi} \int P \, d\omega \left[ -\text{Im} \, \varepsilon(\omega) \right] |E_0(\omega)|^2$ [81], we obtain the single-mode linewidth formula (3). From (29), it is evident that the ST, Petermann, bad-cavity, and incomplete-inversion factors are all included in the term $\frac{P}{a_0^2}$ and generally cannot be separated into the traditional factors of (2) [14].

When the nonlinear coupling coefficient is complex (i.e., when $B \neq 0$), the resonance peak is not only broadened but is also shifted [11]. The shift in center frequency is found by keeping second-order terms in $\delta/a_0$ and calculating the average phase shift:

$$\delta a = \langle \phi \rangle = -\frac{RB}{4a_0^2}.$$  

An identical formula was derived in [29] in a phenomenological instantaneous model.

Figure 2 shows the spectrum of the instantaneous model, which is obtained by numerically solving (15) using a stochastic Euler scheme [82]. Introducing the notation $F(a) \equiv C(a_0^2 - |a|^2) a$ and discretizing time as $a(n + \Delta t) \approx a_n$, the Euler update equation for the nth step is

$$a_n = a_{n-1} + F(a_{n-1}) \Delta t + \sqrt{R} \Delta t \xi,$$

where $\xi$ is a Gaussian random variable of mean 0 and variance 1; i.e., $\xi \in \mathcal{N}(0,1)$. [For the data presented in Fig. 2, $\Delta t$ was decreased until the simulation results converged. In later sections (Fig. 3), we implemented a fourth-order Runge-Kutta method in order to achieve convergence]. The simulated spectra (noisy colorful curves) match the predicted Lorentzian line shapes (solid black curves), which are calculated using (29) and (31). As $\tilde{a}$ increases, the linewidths are broadened and the center frequencies are shifted.

**B. Time-delayed multimode model**

We now turn to the laser spectrum produced by the time-delayed model, where the nonlinearity is dependent on the modal amplitudes at previous times. Although we calculate the linewidth of the full time-delayed N-SALT TCMT equations (11) in Appendix B, we begin this section by considering the simplified case of a spatially homogeneous medium $\gamma(x) \approx \gamma_0$ (this is a good approximation for a uniformly pumped class B laser operating near threshold). In this case, the single-mode time-delayed model takes the form

$$\dot{a} = C \left\{ \gamma_0 \int t \, e^{-|a(t')|^2} \right\} a + f,$$

where $C = \int c(x) \, dx$ is the integrated nonlinear coupling and $c(x)$ is defined in (12). This integro-differential equation can be turned into a first-order ODE by using the modal expansion from Sec. V A: $a = (a_0 + \delta e^{\phi})$, keeping terms to first order in $\delta/a_0$, and introducing the variable

$$\xi(t) = \frac{\gamma_0}{\int t} \int t' e^{-|a(t')|^2} \delta(t').$$

Then (33) and (34) can be recast in the form $\dot{v} = \mathbb{K} v + f$, where $v = [\delta, a_0, \phi, \xi]$. However, most generally, the spatial dependence of $\gamma(x)$ cannot be neglected. The time-averaged deviation $\xi(x, t)$ is therefore spatially dependent, and one obtains an infinite-dimensional problem. To simplify the algebra, we discretize space [e.g., discretizing (11) into a Riemann sum over subvolumes $V_k$] and recover the continuum limit at the end. This yields the discrete-space multimode model:

$$\dot{\tilde{a}}_\mu = \sum_k C_{\mu\nu} \left[ \gamma_k \int t' e^{-|\tilde{a}(t')|^2} \right] a_\nu + f_\mu,$$

where the discretized nonlinear coupling coefficients are $C_{\mu\nu} = \int \gamma_k \, dx c_{\mu\nu}(x)$ (so that $C_{\mu\nu} = \sum_k C_{\mu\nu}^k$), $\gamma_k$ is the relaxation rate at the kth spatial point, and $a_\nu$ is the steady-state amplitude of mode $\nu$.

In Appendix B, we study the statistical properties of the solutions to (35). We introduce the the M-dimensional vectors whose entries are $\Phi_\mu = a_0 \phi \theta_\mu$ (where $M$ is the number of active lasing modes) and we calculate the covariance matrix $\langle \Phi_\mu(t) \Phi_\nu(0) \rangle$. We find that the result is independent of the relaxation rates $\gamma_k$ or the discretization scheme:

$$\langle \Phi(t) \Phi^T(0) \rangle = \begin{bmatrix} \mathbb{R}^2 & \mathbb{R}^2 & \mathbb{R}^2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

The matrices $\mathbb{A}$ and $\mathbb{B}$ correspond to the real and imaginary parts of the coupling matrices, with entries $A_{\mu\nu} = 2a_\mu a_\nu \text{Re}[C_{\mu\nu}]$ and $B_{\mu\nu} = 2a_\mu a_\nu \text{Im}[C_{\mu\nu}]$. $\mathbb{R}$ is the autocorrelation function of the Langevin force vector $f$ [defined in (21)]. The diagonal of this matrix, divided by $|t|$ and by the squared modal amplitude, gives the generalized linewidths

$$\Gamma_\mu = \frac{1}{2\bar{a}_\mu} \{ R_{\mu\mu} + \mathbb{R}^2 \mathbb{B}^{-1} \mathbb{R}^2 (\mathbb{B}^{-1})^T \}_{\mu\mu}.$$  

Therefore, the generalized $\alpha$ factor (which is responsible for linewidth enhancement due to coupling of amplitude and phase fluctuations) is given by

$$\bar{\alpha}_\mu = \frac{1}{R_{\mu\mu}} \{ \mathbb{B}^{-1} \mathbb{R}^2 (\mathbb{B}^{-1})^T \}_{\mu\mu}.$$  

In the single-mode case ($M = 1$), this matrix formula reduces to the single-mode linewidth of the instantaneous model: $2f_{\text{RO}} = \beta(1 + \xi^2)$ [(29) and (30) in Sec. VA].

The linewidth in the time-delayed (class B) model is precisely the same (neglecting side peaks) as in the instantaneous (class A) model. While this result was derived for single-mode class B semiconductor lasers using a phenomenological rate-equation framework [38], we prove that this is generally the case in the multimode inhomogeneous regime. Naively, one might expect to obtain different linewidths due to the longer time over which the fluctuations can grow. However, in Appendix B we obtain a linewidth expression which is independent of the relaxation-oscillation dynamics, which demonstrates that there is a cancellation of two competing processes: As $\gamma_0$ decreases, amplitude fluctuations grow, but they are also averaged over longer periods of time so that their effect is smaller.

Figure 3 presents the simulated spectrum of the time-delayed model in the homogeneous-$\gamma$ limit, which is obtained by numerically integrating (33) (by applying a stochastic Euler scheme, as in Fig. 2). The width of the central peak of the spectrum matches our prediction (29), independent of the value of $\gamma_0$. At intermediate relaxation rates, we also observe side peaks in the spectrum due to amplitude relaxation oscillations (ROs), in addition to the central peak.

C. Side peaks in the time-delayed model

In class B lasers, amplitude fluctuations relax to steady state via relaxation oscillations [45] and, consequently, give rise to side peaks in the spectrum, in analogy with amplitude modulation of harmonic signals. Mathematically, the oscillation arises from the second-order ODE generated by coupling of the $\delta$ and $\xi$ equations (33) and (34), producing the coupled amplitude-gain oscillations. Using the same methods that we applied to calculate the linewidth of the central resonance peak (37), we also calculated the full side-peak spectrum in the multimode regime. Our formula is derived under the fairly general assumption that the central resonance peaks are narrower than the side peaks, which is the relevant regime for many lasers [38]. Although the derivation uses the same techniques as in Appendix B, it is fairly involved and will be provided in a subsequent paper [83]; we only summarize here.

As was shown in Sec. III, far above threshold, the atomic relaxation rate (13) is enhanced and can even be dominated by the electromagnetic field. This modified relaxation rate, and in particular its spatial dependence due to hole-burning effects, has important implications on the RO spectrum. For simplicity, we focus here on the case of $\alpha = 0$. (Note that $\alpha$ factor effects on the RO spectrum have been observed and analyzed using a phenomenological homogeneous time-delayed model in [38].)

In order to see how one can obtain a closed-form expression for the RO spectrum, recall that when calculating the spectrum of the central resonance peak in Sec. VA, we neglected direct amplitude-fluctuation contributions in (26); i.e., in passing from the first to second step, we omitted a term of the form

$$\langle \delta(t)\delta(0) \rangle e^{-i(\phi(t) - \phi(0))}. \tag{39}$$

Adding this term in (26), one finds that the full spectrum consists of an additional term, which is given by the convolution of the real-amplitude fluctuation spectrum $\langle \delta(t)\delta(0) \rangle$ and the spectrum of the central resonance peak. In the instantaneous model, the amplitude autocorrelation function $\langle \delta(t)\delta(0) \rangle$ decays exponentially in time [see (27)] and the omitted term results in near-constant background noise. However, in the time-delayed model, this neglected term is responsible for the RO side peaks.

For simplicity, consider first a model which can be solved straightforwardly; the single-mode homogeneous-$\gamma$ time-delayed model [i.e., $\gamma(x) \approx \gamma_0$ and $\int dx c(x) = C$ as in (33)], which describes uniformly pumped single-mode lasers near threshold. Following the discussion in Sec. VB, we can rewrite (33) and (34) as a set of linear equations and solve for $\langle \delta(t) \rangle$,

$$\delta(t) = \int dt' e^{-\frac{\gamma}{2}(t-t')} \left[ \cosh \left( \frac{\Delta}{2}(t-t') \right) \right] \frac{\gamma_0}{\Delta} \sinh \left( \frac{\Delta}{2}(t-t') \right) f_{\text{RO}}(t') \tag{40}$$

where $\Delta \equiv \sqrt{\gamma_0^2 - 4A^2}$.

In the limit of well-resolved side peaks (e.g., $\kappa \ll \gamma_0 \ll A$), the amplitude autocorrelation function is approximately

$$\langle \delta(t)\delta(0) \rangle \approx \frac{R}{2} \left[ \frac{\sin \sqrt{\gamma_0 A} t}{\sqrt{\gamma_0 A}} + \frac{\cos \sqrt{\gamma_0 A} t}{\gamma_0} \right] e^{-\frac{\gamma_0 t}{2}}. \tag{41}$$

Thus, additional peaks in the spectrum arise at frequencies $\omega_{\text{RO}} = \omega_0 \pm \sqrt{\gamma_0 A}$ with widths $\gamma_0$. In the high-$Q$ limit near threshold, $A$ is proportional to the cavity decay rate $\kappa$, giving the expected behavior for the RO frequency. The side-peak amplitudes $\frac{R}{2} \left[ \frac{1}{\sqrt{\gamma_0 A}} \right] + \frac{1}{\gamma_0}$ diverge in the limit of $\gamma_0 \to 0$ (that is, when amplitude fluctuations are not small compared to the steady-state mode amplitude), but this is also the regime in which our analysis of the spectrum (Secs. VA and VB) breaks down. The inset in Fig. 4(b) shows the simulated spectrum of the homogeneous time-delayed model (33) (the same data was also shown in Fig. 3, but we include here the theoretical formula for the side-peak spectrum). The exact numerical solution of (33) (blue curve) reproduces the analytic spectrum prediction of (42) (red curve).

In the limits of extremely small or large relaxation rates $\gamma_0$ (compared to $A$), the side peaks disappear. In the former limit, they merge with the central resonance peak and in the latter case, they merge with the background noise. This
behavior can be explained by inspection of the $\dot{\delta}$ and $\dot{\xi}$ equations (33) and (34) in the appropriate limits. When the relaxation rate is very large, the time-delayed model reduces to the instantaneous model, which represents the case where the atomic population follows the field adiabatically. In the opposite limit of extremely small relaxation, the field follows the atomic population adiabatically. In other words, a clear separation of atomic and optical time scales will result in the absence of RO side peaks.

In the most general spatially inhomogeneous time-delayed model, the full spectrum takes the simple form

$$S_\alpha(\omega) = \frac{\Gamma}{\omega^2 + \left(\frac{\Gamma}{T}\right)^2} + \frac{\Gamma}{\omega^2 \left(1 - \int dx \frac{A(x)\gamma(x)}{\omega^2 + (\Gamma x + \gamma(x))^2} + \int dx \frac{A(x)\gamma(x)\gamma(x) + 1}{\omega^2 + (\Gamma x + \gamma(x))^2}\right)^2},$$

(42)

where $A(x)$ is the real part of the local nonlinear coupling [defined in (23)], $\gamma(x)$ is the effective decay rate, and $\Gamma$ is the central-peak linewidth. (This formula is valid when the central resonance peak is narrower than the side peaks $\Gamma \ll \gamma_1$. Like our linewidth formula, this formula is easy to evaluate via spatial integrals of the SALT solutions.

While the homogeneous time-delayed model near threshold agrees with standard results on relaxation oscillations [38], the full model above threshold, combined with SALT, is able to include effects not contained in other treatments. As the pump is increased far above threshold, the effects of stimulated emission strongly increase the atomic relaxation rate, and spatial hole burning causes that rate $\gamma(x)$ to vary substantially in space [see (13)]. These two effects cause both a shift and a broadening of the side peaks compared to the near-threshold result. Figure 4 shows the dressed decay rate $\gamma(x)$ and the side-peak spectrum $S_{RO}(\omega)$ [as given by the second term of (42)], based on a SALT calculation of a $1$D PhC laser, at four different pump values well above threshold. [The pump value is controlled via the parameter $D_p$ in (A3), and we denote the threshold value of $D_p$ by $D_{th}$.] This type of cavity [depicted in the inset of Fig. 4(a)] supports a single mode at the simulated parameter regime, which is localized near the defect region. (Further discussion of this structure is given in Sec. VI.A.) As can be seen from Fig. 4(a), the decay rate $\gamma(x)$ is enhanced at high-intensity regions (i.e., near the defect), and it increases further as the pump increases. Figure 4(b) demonstrates the shifting and broadening of the side peaks.

VI. THE GENERALIZED $\alpha$ FACTOR

Our TCMT derivation of the linewidth formula yields a generalized $\alpha$ factor (38) which depends on the eigenmodes $E_\alpha(x)$ and eigenfrequencies $\omega_\alpha$ of the full nonlinear SALT equations. This is an advance over previous linewidth formulas; the $ab$ initio scattering-matrix linewidth formulas did not obtain an $\alpha$ factor [13,14], whereas other traditional laser theories that derived $\alpha$ factors could not handle the full nonlinear equations [50]. Therefore, in the following section, we focus on the generalized $\alpha$ factor. We compare the generalized and traditional factors in Sec. VI.A, and then we evaluate these factors in the single-mode (Sec. VI.B) and multimode (Sec. VI.C) regimes.

A. Comparison with traditional $\alpha$ factor

Linewidth broadening due to amplitude-phase coupling (that is, the $\alpha$ factor linewidth enhancement) was first studied in the 1960s by Lax in the context of single-mode detuned gas lasers [9]. The Lax $\alpha$ factor is $1 + \alpha L^2$, where $\alpha_0$ is the normalized detuning of the lasing frequency from the atomic resonance, i.e., $\alpha_0 = \frac{\hbar\nu_0 - \omega_0}{\gamma_1}$, which is equal to the ratio of the real part of the gain permittivity to its imaginary part, or equivalently the ratio $\frac{\text{Re} \Delta n}{\text{Im} \Delta n}$, where $\Delta n$ is the refractive-index change due to fluctuations in the gain. Two decades later, Henry derived an amplitude-phase coupling enhancement factor of the same general type in semiconductor lasers [11], $\alpha_0 = \frac{\text{Re} \Delta n}{\text{Im} \Delta n}$, but in the latter case these refractive-index changes arise from carrier-density fluctuations and take a different form. Here we are considering atomic gain media, so our $\alpha$ factor generalizes the Lax form.

The difference between our single-mode generalized $\alpha$ factor (30) and that of Lax arises because we take into account spatial variation in the gain permittivity due to spatial hole burning and also the non-Hermitian (complex) nature of the lasing mode. Hence, we expect our factor to reduce to the Lax factor in some limits. For instance, consider the situation that was discussed in the last paragraph of Sec. II of a low-loss $1$D Fabry-Pérot cavity laser, operating near threshold. In this case, the nonlinear coupling coefficient is approximately $C \approx \frac{-\alpha_0}{2\varepsilon} \int dx \frac{E_\alpha^2}{E_\alpha^2}$, and one can show that the generalized $\alpha$ factor is $\bar{\alpha} = \frac{\text{Im} C}{\text{Re} C} \approx \frac{\text{Re} \Delta n}{\text{Im} \Delta n}$ (the last approximation is valid since in
essentially all realistic cavities, the modes can be chosen to be predominantly real, i.e., have small imaginary parts).

In many cases, however, our \( \tilde{\alpha} \) deviates from the traditional factor \( \alpha_0 \). An obvious example is when the lasing frequency precisely coincides with the atomic resonance frequency. In this case, the traditional factor vanishes, but \( \tilde{\alpha} \) does not necessarily vanish. In the next section, we calculate and discuss the characteristic properties of the generalized \( \alpha \) factor for two 1D laser structures.

### B. Generalized single-mode \( \alpha \) factor

In this section, we evaluate the differences between the generalized and traditional \( \alpha \) factors in 1D model systems. We solve the full nonlinear SALT equations using our recent finite-difference frequency-domain (FDFD) SALT solver [21].

The generalized factor \( \tilde{\alpha} \) can deviate significantly from the traditional factor \( \alpha_0 \) when the latter is large (a similar argument was made in [44]). To see this, let us write the nonlinear coupling coefficient qualitatively as \( C \propto (1 + i\alpha_0)(1 + i\beta) \), where the term \( 1 + i\alpha_0 \) is associated with the atomic line shape \( \frac{\alpha}{\alpha_0 - \omega_0 + i\gamma} \), and the term \( 1 + i\beta \) is a complex factor due to the remaining integral factors (we refer to the latter term as the modal contribution to the \( \alpha \) factor). Typically, \( \beta \ll 1 \) and, consequently, the generalized factor is approximately \( \tilde{\alpha} \approx \alpha_0 + \beta(1 + \alpha_0^2) \), so the difference between the generalized and traditional factors grows quadratically with \( \alpha_0 \).

To verify this argument, we study a model system in which the magnitude of \( \alpha_0 \) can be controlled. Consider a quarter-wave dielectric PhC, with a defect at the center of the structure [the geometry is depicted in the upper inset of Fig. 5(a), similar to the structure that was studied in Fig. 4]. Adding enough layers of the periodic structure on each side of the defect to mimic an infinite structure, one finds that the system has a localized mode in the vicinity of the defect (lower inset), whose resonance frequency is fixed to a real value within the energy gap [33]. To study finite-threshold lasers, we introduce gain and some passive loss (i.e., a positive imaginary permittivity term, which pushes the resonance poles away from the real axis in the complex plane). Since the resonance frequency of the defect mode is fixed by the geometry, by varying the resonance frequency of the gain, we control the detuning of the lasing mode from the atomic resonance, thus controlling the size of \( \alpha_0 \). As demonstrated in the figure, the deviation \( |\tilde{\alpha} - \alpha_0| \) grows as the detuning \( \Delta \nu \equiv \frac{\omega_0 - \omega_\nu}{\hbar} \) increases.

The openness of the cavity also results in an enhancement of the \( \alpha \) factor; the more open it is, the larger is the necessary imaginary part of the lasing mode, which causes a deviation from the standard formula. In order to test this prediction, we evaluate the generalized \( \alpha \) factor for an open-cavity laser [Fig. 5(b)], where we can control the radiative loss rate through the cavity walls and, consequently, this part of the modal contribution to \( \tilde{\alpha} \). We consider a cavity which consists of a dielectric slab (with permittivity \( \varepsilon_c \)) surrounded by air on both sides, with gain spread homogeneously inside the slab (upper rightmost inset). The reflectivity of the cavity walls is determined by the difference in cavity and air permittivities \( \Delta \varepsilon = \varepsilon_c - \varepsilon_0 \). For relatively small dielectric mismatch, the cavity is relatively low-\( Q \) and our \( \alpha \) factor differs significantly from the Lax factor. As \( \Delta \varepsilon \) increases and the cavity \( Q \) increases, the generalized \( \alpha \) factor converges to the original factor, so that the red and blue curves in the figure overlap.

Unlike a PhC defect-mode cavity where there is a finite bandwidth of confinement [33], this dielectric cavity has an infinite number of possible lasing resonances and thus when we sweep \( \Delta \varepsilon \), the \( \alpha \) factor peaks periodically. This is because the free spectral range of the cavity is \( \Delta \nu \approx \frac{2\pi c}{\Delta \varepsilon} \frac{1}{\varepsilon_c} \) [84] and, therefore, changing \( \varepsilon_c \) corresponds to shifting the passive resonances and, consequently, the lasing modes. Every time a lasing mode crosses an atomic resonance, \( \alpha_0 \) vanishes and, correspondingly, \( \tilde{\alpha} \) becomes very small. The traditional factor is maximized when the atomic resonance is equidistant from two passive modes. The peak value is proportional to the free spectral range and, therefore, we find that it is proportional to \( 1/\sqrt{\varepsilon_c} \). This type of effect may not have been observed previously because in macroscopic cavities, the cavity resonances are very dense on the scale of the gain bandwidth, so the lasing mode can never be substantially detuned. However, in microcavities with large free spectral range, this could be an important effect. Another intriguing property of the generalized \( \alpha \) factor is that it varies

![FIG. 5. (Color online) (a) The generalized (blue) and traditional (red) \( \alpha \) factors of a PhC laser vs relative detuning \( \Delta \nu \equiv \frac{\omega_0 - \omega_\nu}{\hbar} \). (Upper inset) Quarter-wave PhC geometry (see caption of Fig. 4). The gain parameters are \( \gamma_c = 3 \text{ mm}^{-1} \) and a varying \( \omega_\nu \). (Lower inset) Intensity distribution of the lasing mode. (b) \( \alpha \) factor for an open-cavity laser vs passive permittivity \( \varepsilon_c \). Blue (red): generalized (traditional) \( \alpha \) factor. (Upper inset) Dielectric slab, of permittivity \( \varepsilon_c \), bounded by air on both sides, containing gain atoms with \( \omega_\nu = 15 \text{ mm}^{-1} \), \( \gamma_c = 3 \text{ mm}^{-1} \). (Lower inset) Intensity distribution of the lasing mode. (Leftmost inset) Enlarged segment of the main plot, around \( \varepsilon_c = 7 \).](image)
discontinuously at the peaks (as is shown more clearly in the upper leftmost inset). The traditional factor $a_0$ depends only on the mode detuning from resonance, so it approaches the same value on different sides of the peak. In contrast, the generalized factor $\tilde{a}$ depends on the mode profile $E_n$, which differs between the two interchanging laser modes on different sides of the peak, producing the observed asymmetry.

**C. Generalized multimode $\alpha$ factor**

Our multimode linewidth formula includes linewidth corrections from neighboring modes, which enter through the generalized $\alpha$ factor (since phase fluctuations in each of the modes couple to amplitude fluctuations in all other modes due to saturation of the gain). According to the traditional ST formula (2), when phase cross correlations between different modes are neglected, each resonance-peak width is inversely proportional to the corresponding modal output power. We find that when phase cross correlations are included, the linewidth of each mode is a sum of inverse output powers of all the other modes, a type of multimode ST relation. To see how this comes about, recall that the generalized $\alpha$ factor, as given by (38), is proportional to $[\text{BA}^{-1}]_{ij} \Re (\text{BA}^{-1})^T_{ij}$, where $a_{0j}$ is the steady-state amplitude of the $j$th mode. Therefore, the multimode $\alpha$ factor is proportional to the sum $\tilde{a}_{0j}^2 \sum_j \langle \sigma_{n_j} \rangle^{-1} a_{n_j}^0$, i.e., a sum over terms which scale as inverse output powers.

In the two-mode case, the linewidth formula for a lasing mode in the presence of a neighboring mode is given explicitly by

$$\Gamma_1 = \frac{R_{11}}{2a_{10}} + \frac{R_{11}}{2a_{10}} \left[ \frac{C_{11}^R C_{22}^R - C_{21}^R C_{12}^R}{C_{11}^R C_{22}^R - C_{12}^R C_{21}^R} \right]^2 + \frac{R_{22}}{2a_{20}} \left[ \frac{C_{11}^R C_{12}^R - C_{12}^R C_{11}^R}{C_{11}^R C_{12}^R - C_{22}^R C_{21}^R} \right]^2,$$

(43)

where $C_{ij}^R \equiv \Re C_{ij}$ and $\tilde{C}_{ij}^R \equiv \Im C_{ij}$. (A similar expression was derived in [39] by using a phenomenological version of the two-mode TCMT equations.) As predicted by the multimode ST relation, the last term in (43) is inversely proportional to the output power of the second mode $a_{20}^2$. This term becomes significant when the power in the first mode greatly exceeds the power in the second mode (i.e., when $P_1 \gg P_2$), correcting the unrealistic ST prediction that the linewidth vanishes when $P_1 \to \infty$; a similar argument was made in [39]. Figure 6(a) presents the spectrum of a two-mode instantaneous model (15) in the parameter regime where cross-correlations between the two modes are significant. The linewidth of the simulated spectrum (red curve) is in complete agreement with the generalized formula (43) (black curve), but deviates substantially from the single-mode formula (29) (blue curve). In order to reach the regime where this deviation is substantial, in practice, one needs to design a cavity in which the two lasing modes have comparable amplitudes and detunings from the atomic resonance frequency.

Equation (43) predicts an unphysical divergence near the second threshold, i.e., when $a_{20} \to 0$ [see black curve in Fig. 6(b)]. In retrospect, this singularity is to be expected, since the assumptions of our derivation break down in this limit. (Note that an equivalent divergence was present in [39].) In calculating the phase variance, we assumed that amplitude fluctuations in all modes were small compared to the steady-state amplitudes ($\delta_1 \ll a_{01}$), and this assumption is no longer valid near threshold. The N-SALT TCMT equations (11) are still valid, however; it is only their analytical solution for $\langle \Phi \Phi^T \rangle$ that is problematic. Therefore, we study the threshold regime numerically, via stochastic simulations of the N-SALT TCMT equations. As shown in Fig. 6(b), the simulated linewidth of the first mode approaches a finite value near the second threshold (red curve), and this value is significantly larger than the linewidth prediction one obtains when neglecting the second mode (blue curve). Even at the threshold, noise in the second mode mixes with the first mode through off-diagonal nonlinear coupling terms, thus increasing the linewidth.

Linewidth enhancement at the thresholds of neighboring lasing modes suggests that the linewidth must also be enhanced.

**FIG. 6.** (Color online) (a) Spectrum near a resonance peak in the presence of an additional mode. Numerical simulations of (15) (red curve) and analytic single-mode (blue) and multimode (black) formulas. The simulation parameters are chosen so that there are two lasing modes with the same steady-state amplitudes $a_{10} = 1$ and diffusion coefficients $R_{11} = 0.05$, and with substantial cross correlations: $C_{1k} = 5$, $C_{ij} = 4 + 4i k \neq l$ (in arbitrary frequency units). (b) Linewidth of central resonance peak vs output power in the neighboring mode [$a_{10} = a_{20} = 1$] is encircled and corresponds to the parameter values of Fig. 6(a).
below the modal thresholds (in the regime where radiation from nonlasing modes is incoherent, commonly called ASE). We believe that this phenomenon could be explored using a future generalization of our formalism, with some modifications (extending earlier work [28,40] on linewidth enhancement from ASE).

VII. FULL-VECTOR 3D EXAMPLE

In order to illustrate the full generality of our approach, we apply it in this section to study a 3D PhC laser. The steady-state properties of this system (i.e., the lasing threshold and mode characteristics) were previously explored in [21]. We use those solutions here to calculate the laser linewidth [using (3)], and we compare the relative contributions of the various correction factors.

The simulated PhC consists of a dielectric slab patterned by a hexagonal lattice of air holes [Fig. 7(a)]. A defect is introduced by decreasing the radii of seven holes at the center of the structure [85], giving rise to a doubly degenerate mode which is situated at the defect (spatially) and in the band gap (spectrally). We select the TE-like mode out of the degenerate pair by imposing even and odd reflection symmetry at $x = 0$ and $y = 0$, respectively, as well as an even reflection symmetry at $z = 0$. Staying close to a potential experimental realization, we choose the pump profile to be uniform inside the high-index dielectric near the defect region, and zero elsewhere. We solve the SALT equations using our scalable FDFD solver and track the evolution of the first lasing mode upon increasing the pump strength from zero to five times the first-threshold value.

Typically, realistic laser structures do not use two-level gain media, but employ a more complex optical scheme which involves multiple levels and transitions in order to achieve significant inversion and depletion of the ground-state population. In this section, we apply our formalism to a four-level gain medium [Fig. 7(b)], using a generalization of SALT [73], which finds the stationary multimode lasing properties of an N-level gain medium. As shown in [73], an N-level system can be mapped into an effective two-level system, which obeys the (two-level) SALT equations with renormalized pump ($D_p$) and atomic relaxation rates ($\gamma_i$). Consequently, the linewidth of a four-level laser will be given by our generalized formula (3) with the appropriately renormalized coefficients. By choosing the decay rate between the lasing transition levels ($\gamma_2$ in Fig. 7(b)) to be much smaller than the decay rates into the upper ($\gamma_3$) and out of the lower ($\gamma_1$) states, we can achieve substantial inversion and ground-state depletion. Consequently, the incomplete-inversion factor is approximately $n_{in} \approx 1$, close to typical measured values [47].

Figure 7(c) presents the traditional and new correction factors (dashed and solid lines, respectively), as defined in Table I. We find that those factors are relatively small for this system and, consequently, the deviations between the new and traditional factors are small. A small Petermann factor arises since the first lasing mode has a relatively high quality factor (i.e., the cold-cavity resonance pole is at $\omega_0 = 1.725 - 0.005 i$ with a quality factor of $Q \approx 700$, in agreement with experimental realization [85]). Moreover, the cold-cavity resonance lies well within the gain bandwidth, resulting in small $\alpha$ and bad-cavity corrections. The generalized factor $\tilde{\alpha}$ (solid purple line) is obtained from from Eqs. (17) and (30). Deviations of $\tilde{\alpha}$ from the traditional factor $\alpha_0 \equiv \frac{\beta_{l|N-1}}{\beta_{l|N-1}}$ (dashed purple line) are due to modal contributions to the $\alpha$ factor (see Sec. VI B). The generalized Petermann factor (solid blue curve) is compared against the traditional factor (dashed blue line), which is expressed in terms of the SALT mode (instead of the passive-cavity mode). The cavity region is taken to be the entire high-index medium. (Note that the generalized and traditional factors agree at threshold.) Both the Petermann and $\alpha$ factors increase the linewidth. However, the generalized and traditional bad-cavity factors (full and dashed red curves, respectively) lead to linewidth reduction.

Last, we evaluate the incomplete-inversion factor $\tilde{n}_{in}$. The inversion $D(x)$ is found from the SALT solutions of the effective two-level system. The excited-state population $N_2(x)$ can be derived straightforwardly, using the results of [73] as follows. Assuming that the populations in the nonlasing levels $|0\rangle$ and $|3\rangle$ are at steady state, one can express those populations in terms of the populations in the lasing transition $|1\rangle$ and $|2\rangle$. Then, by invoking the density conservation condition, $\sum_j N_j = n$, where $n$ is the atom number density and $N_j$ are the individual level populations, one finds that the population...
in (2) is given by
\[ N_2 = \frac{n + \tau D}{1 + \tau}, \tag{44} \]
where \( \tau \equiv 1 + \frac{2n}{D} + \frac{2i}{D} \). Having obtained expressions for \( D \) and for \( N_2 \), we have all that is needed to calculate the incomplete-inversion factor \( \tilde{n}_{p,g} \). We define the "linear incomplete-inversion factor" \( (n_{p,g} \text{ dashed green line}) \) as the ratio \( \frac{N_1(D_p)}{D_p} \) [i.e., both the excited-state population (44) and the inversion are evaluated at \( D = D_p \), neglecting hole-burning effects]. The "nonlinear incomplete-inversion factor" \( (\tilde{n}_{p,g} \text{ solid green line}) \) is defined in Table I. The nonlinear factor \( \tilde{n}_{p,g} \) coincides with the linear factor \( n_{p,g} \) at threshold, but exceeds the traditional factor at higher pumps. We also plot the total linewidth correction, which is defined as the product of the (traditional and new) Petermann, \( \alpha \), bad-cavity, and incomplete-inversion factors.

VIII. CONCLUDING REMARKS

We presented a generalized multimode linewidth formula, obtained from the N-SALT TCMT equations for the lasing-mode amplitudes, which we derived starting from the Maxwell-Bloch equations and using the fluctuation-dissipation theorem to determine the statistical properties of the noise. Our generalized linewidth formula (3) reduces to the traditional formula (2) for low-loss cavities and simple lasing structures, but deviates significantly from the traditional theories for high-loss wavelength-scale laser cavities. By basing our derivation on the SALT steady-state lasing modes, it is possible to apply our formula to cavities of arbitrarily complex geometry (e.g., PhC or microdisk lasers [15–18]) and arbitrary openness (e.g., random lasers [43]). Also, since SALT includes high accuracy the effects of spatial hole burning, our formula includes both gain saturation and the spatial variation of the gain permittivity well above threshold, plus all effects due to modal couplings. From a computational point of view it is important to point out that our formula is analytical and can be evaluated immediately from the output of a numerical SALT calculation without any significant computational effort. A paper describing a brute-force numerical validation of our theory against numerical solution of the Maxwell-Bloch equations is currently being prepared [41]. Given only the laser geometry, the pumping profile, and characteristic properties of the gain (i.e., its resonance frequency \( \omega_p \) and decay rate \( \gamma_2 \)), our formula enables linewidth calculation, including a generalized \( \alpha \) factor and accounting for temperature variations, at a level of generality that was not possible before. This generality is most important, of course, in cases where the new result is substantially different than previous theories, and it would be interesting to study laser cavities in which the discrepancy is as large as possible.

One such case is that of lasers which contain exceptional points (EPs) in their spectrum, which are points of degeneracy where two (or more) eigenfrequencies and eigenfunctions coalesce [65,86]. EPs in laser systems have been explored recently, both theoretically [87] and experimentally [88]. At the EP, the modes become self-orthogonal and that causes the denominator of (3) to vanish and is already known to greatly enhance the Petermann factor [89]. Since a similar denominator appears in the integrals defining our generalized \( \alpha \) factor (12) and (30), we expect that \( \tilde{\alpha} \) will differ substantially from previous results near an EP (and similarly for the inhomogeneous-temperature correction).

An important and exciting addition to the theory would be a treatment of ASE from modes below threshold; we believe this can be achieved by deriving TCMT equations for below-threshold (passive) modes, in which there is no steady-state oscillation (generalizing previous ASE work which used simplified models [28, 40]). Incorporating the ASE contribution to the spectrum will allow us to follow the noise through the lasing thresholds, correcting the unphysical divergence which was discussed in Sec. VII B. More importantly, treating below-threshold ASE should allow an ab initio theory of light emitting diodes (LEDs) in arbitrary cavities.

Future work could also incorporate several additional corrections that were not treated in this paper. Our derivation applies to isotropic materials described by a scalar permittivity \( \varepsilon \), but extension to anisotropic permittivity \( \hat{\varepsilon} \), magnetic permeability \( \mu \), and even bianisotropic materials would be very straightforward (e.g., for an anisotropic \( \hat{\varepsilon} \), the only change is that \( \varepsilon F^2 \) factors and similar are replaced by \( \varepsilon F \cdot \langle \hat{\varepsilon} E \rangle \), etc., as in [55]). As discussed in Sec. VC, we are also able to exploit our framework to analytically solve for the relaxation-oscillation side-peak spectra and are currently preparing a paper presenting this analysis [83]. We believe it will be possible to extend our formalism to handle non-Lorentzian line shapes arising from frequency dependence (correlations) in the noise within the laser linewidth [60–64], as also discussed in Sec. IV. Instead of treating the noise spectrum \( S_F(\omega) \) as a constant \( S_F(\omega_0) \), one needs to include a first-order correction, e.g., by Taylor expanding \( S_F(\omega) \) around \( \omega_0 \); it might be convenient to fit \( S_F(\omega) \) to a Lorentzian matching the amplitude and slope at \( \omega_0 \), since the Fourier transform of a Lorentzian is an exponential that should be easy to integrate. Finally, as noted above, although our derivation was for the two-level Maxwell-Bloch equations, a similar approach should apply to more complex gain media (including multilevel atoms [73], multiple lasing transitions, and gain diffusion [74]). The N-SALT linewidth theory can be generalized to account for these laser models following along the lines of our approach here.

ACKNOWLEDGMENTS

This work was partially supported by the Army Research Office through the Institute for Soldier Nanotechnologies under Contract No. W911NF-13-D-0001. A.D.S. and A.C. acknowledge the support of NSF Grant No. DMR-1307632. C.Y.D. acknowledges the support of Singapore NRF Grant No. NRF2FF2012-02. The authors would like to thank Bo Zhen, Aristeidis Karalis, Amir Rix, Owen Miller, and Homer Reid for helpful discussions.

APPENDIX A: DERIVATION OF N-SALT TCMT

In this Appendix, we derive the TCMT equations for the lasing-mode amplitudes. Our starting point is the Maxwell-Bloch equations [2,22], which describe the dynamics of the...
electromagnetic field in a resonator interacting with a two-level gain medium,
\[ \mathbf{V} \times \nabla \times \mathbf{E} + \varepsilon_c \mathbf{E} = -4\pi \mathbf{P} + \mathbf{F}_S, \]  
(A1)
The steady-state inversion \( D_0(a_0) \) (which is a notation shortcut for \( D_0(\{E_\mu\}_\mu,\{\omega_\mu\}_\mu,\{a_\mu\}_\mu) \)) is given by
\[ D_0(a_0) = \frac{D_p}{1 + \sum_\mu \frac{\gamma_{\pm}}{2(\omega_\mu - \omega_0 + i\gamma_\parallel )} |a_\mu|^2 |E_\mu|^2}. \]  
(A8)
To avoid possible confusion, note that in previous SALT works, the steady-state inversion was denoted by \( D \) and \( D_0 \) was the external pump parameter, whereas in this work, \( D_0 \) is the steady-state inversion and \( D \) is the external pump parameter.

2. Noise-driven Maxwell-Bloch equations
In the presence of a small noise source, the electric field and polarization can be written as superpositions of the steady-state lasing modes with time-dependent amplitudes \( a_\mu(t) \) and \( b_\mu(t) \):
\[ \mathbf{E}(\mathbf{x},t) = \sum_\mu \mathbf{E}_\mu(\mathbf{x}) a_\mu(t) e^{-i\omega_\mu t}, \]
\[ \mathbf{P}(\mathbf{x},t) = \sum_\mu \mathbf{P}_\mu(\mathbf{x}) b_\mu(t) e^{-i\omega_\mu t}. \]  
(A9)
Substituting the perturbation ansatz (A9) into the polarization equation (A2), we obtain
\[ (\dot{a}_\mu + i\omega_\mu b_\mu) \mathbf{P}_\mu = -i(\omega_\mu - i\gamma_\parallel) b_\mu D_0(a_0) \]  
\[ \frac{i}{4\pi} \frac{\gamma_\parallel}{\omega - \omega_\mu + i\gamma_\parallel} \mathbf{P}_\mu D_0 \]  
(A10)
Taking the Fourier transform and rearranging terms, we find
\[ \hat{D}_\mu \mathbf{P}_\mu = \frac{1}{4\pi} \frac{\gamma_\parallel}{\omega - \omega_\mu + i\gamma_\parallel} \hat{a}_\mu \ast \hat{D} \mathbf{E}_\mu. \]  
(A11)
where we have introduced the shifted frequency \( \omega \equiv \omega_\mu + \Omega \) and the Fourier-domain envelopes \( \hat{\mathbf{E}}_\mu(\Omega) = \hat{\mathbf{a}}_\mu(\omega - \omega_\mu), \hat{\mathbf{B}}_\mu(\Omega), \) and \( \hat{D}(\Omega) \). The asterisk * denotes a convolution.

Next, consider Eq. (A1) in the frequency domain
\[ \mathbf{V} \times \nabla \times \mathbf{E} - \omega^2 \varepsilon_c \mathbf{E} + 4\pi \mathbf{P} = \mathbf{F}_S. \]  
(A12)
When the spacing between adjacent lasing modes is much larger than their linewidths, a noise source with frequency \( \omega \approx \omega_\mu \) excites only the mode \( E_\mu(x) \). Equivalently, the Green’s function can be approximated by the contribution of the single pole at \( \omega_\mu \). (Note that we require only that the peaks in the laser spectrum above threshold are nonoverlapping; we do not require isolated resonances in the passive-cavity spectrum.) Therefore, at frequencies \( \omega \approx \omega_\mu \), we can substitute (A11) into (A12) and obtain an effective equation for the noise-driven field \( \mathbf{E}_\mu(x,\omega) \):
\[ \mathbf{V} \times \nabla \times \mathbf{E}_\mu(x,\omega) = \mathbf{F}_S(x,\omega), \]  
(A13)
where the effective permittivity \( \varepsilon(\omega,a) \) is given by
\[ \varepsilon(\omega,a) \mathbf{E}_\mu(x,\omega) = \varepsilon_c \hat{a}_\mu + \frac{\gamma_\parallel}{\omega - \omega_\mu + i\gamma_\parallel} \hat{D} \ast \hat{a}_\mu \mathbf{E}_\mu(x). \]  
(A14)
The second variable of \( \varepsilon(\omega,a) \) denotes the implicit dependence of \( \varepsilon \) on the modal amplitudes \( a_\mu \) through the Fourier transform of the inversion \( \hat{D} \). We calculate \( \hat{D} \) explicitly in the next section.
3. The atomic inversion

The noise source $\hat{F}_S$ perturbs the modal amplitudes $a_n$ from steady state, causing a change in the atomic inversion $D$. We neglect dispersion corrections to $D$ (which amounts to setting $b_{\nu} = 0$ in (A10) [37]) as these corrections do not affect the linewidth formula to leading order in the noise [see discussion following (9) in the main text]. From (A3) and (A10), we obtain

$$D = -\gamma_{\nu} \left[ D - D_0 + \sum_{\nu} \frac{\gamma_{\nu}^2}{(\omega_{\nu} - \omega_{0})^2 + \gamma_{\nu}^2} |a_n|_2^2 D |E_n|^2 \right].$$

(A15)

In order to solve (A15), we linearize the time-dependent products $|a_n|^2 D$ in the sum around the steady state $|a_n|^2 D \approx a_n^0 D_0 + D_0 (|a_n|^2 - a_n^2) + a_n^2 (D - D_0)$, where $D_0$ is the steady-state (SALT) inversion (A8). To simplify the notation, we define the local decay rate

$$\gamma(x) \equiv \gamma_{\nu} \left[ 1 + \sum_{\nu} \frac{\gamma_{\nu}^2}{(\omega_{\nu} - \omega_{0})^2 + \gamma_{\nu}^2} |a_n|^2 |E_n|^2 \right].$$

(A16)

The second term in (A16) gives precisely the increased atomic decay rate due to stimulated emission. Using the definitions above, (A15) becomes

$$\dot{D} = -\gamma(x)(D - D_0) - \gamma_{0} D_0 \sum_{\nu} \frac{\gamma_{\nu}^2}{(\omega_{\nu} - \omega_{0})^2 + \gamma_{\nu}^2} |E_n|^2 (|a_n|^2 - |a_n|^2),$$

(A17)

which we can integrate, and obtain

$$D = D_0 + \sum_{\nu} D_0 \left[ \frac{\gamma_{\nu}^2}{(\omega_{\nu} - \omega_{0})^2 + \gamma_{\nu}^2} |E_n|^2 \right] \times \gamma_{\nu} \int dt \, e^{-\gamma_{\nu}(x)|a_n|^2 - |a_n(t')|^2}. $$

(A18)

Having derived an explicit expression for $D(t)$, we substitute its Fourier transform $\hat{D}$ into the effective permittivity expansion (A14) and obtain

$$\epsilon(\omega, a) \hat{E}_\mu \approx \epsilon(\omega, a_0) \hat{E}_\mu + \sum \chi_{\nu}(\omega, a_0) \Delta a_{\nu} \ast \hat{E}_\mu,$$

(A19)

where $\epsilon(\omega, a_0)$ is the steady-state SALT permittivity which was defined in (A7), $\chi_{\nu}(\omega, a_0)$ is the permittivity differential due to deviation in the modal amplitude $a_n$ [which we denote by $\frac{\delta a_n}{\delta a_0}$ in the text, e.g., in (12)],

$$\chi_{\nu} \equiv \frac{\gamma_{\nu}}{\omega - \omega_{0} + i \gamma_{\nu} D_0 \left[ \frac{\gamma_{\nu}^2}{(\omega_{\nu} - \omega_{0})^2 + \gamma_{\nu}^2} |E_n|^2 \right] \gamma_{\nu}}.$$  

(A20)

and $\Delta a_{\nu}$ is the Fourier transform of the time-averaged modal deviation from steady state,

$$\Delta a_{\nu} = \gamma(x) \int dt' e^{-\gamma_{\nu}(x)|a_n|^2 - |a_n(t')|^2}.$$  

(A21)

Substituting the permittivity expansion (A19) into Maxwell’s equation (A13), we obtain

$$[\nabla \times \nabla \times - \omega^2 \epsilon(\omega, a_0)] \hat{E}_\mu(x, \omega) = \hat{F}_{NL}(x, \omega) + \hat{F}_S(x, \omega),$$

(A22)

where the nonlinear restoring force is

$$\hat{F}_{NL}(x, \omega) = \omega^2 \sum_{\nu} \chi_{\nu}(\omega, a_0) \Delta a_{\nu} \ast \hat{E}_\nu(x, \omega).$$

(A23)

The left-hand side of (A22) is just the linearized steady-state equation (A6), and the nonlinear correction to the effective permittivity due to the noise $\hat{F}_S$ appears as an additional source term $\hat{F}_{NL}$. As noted above, the noise-driven field $\hat{E}_{\nu}$ is found by integrating the Green’s function of the steady-state equation (A6) over the noise terms $\hat{F}_{NL}$ and $\hat{F}_S$. In the following section we derive an approximate formula for the Green’s function.

4. The linearized steady-state Green’s function

The single-pole approximation of the Green’s function is valid for frequencies near the resonances $\omega \approx \omega_\mu$, as long as the spectrum consists of nonoverlapping resonance peaks, i.e., when the spacing between resonant modes exceeds the modal linewidths. First, let us rewrite the left-hand side of (A6) as an operator $\mathcal{L}_\omega$ acting on the field $\hat{E}(x, \omega)$:

$$\mathcal{L}_\omega \hat{E}(x, \omega) = \frac{\hat{F}_{NL}(x, \omega)}{\omega^2}.$$  

(A24)

Next we choose a complete set (see below) of eigenfunctions $\hat{E}_n(x, \omega)$ and eigenvalues $\lambda_n(\omega)$ of the operator $\mathcal{L}_\omega$:

$$\hat{L}_\omega \hat{E}_n(x, \omega) = \lambda_n(\omega) \hat{E}_n(x, \omega).$$

(A25)

We define the inner product of two vector fields, $\mathbf{A}(x)$ and $\mathbf{B}(x)$, as $(A, B) \equiv \int dx \mathbf{A}(x) \cdot \mathbf{B}(x)$. The operator $\mathcal{L}_\omega$ is complex symmetric under this inner product, i.e., $(A, \mathcal{L}_\omega B) = (\mathcal{L}_\omega A, B)$ [33,65]. Therefore, we use unconjugated inner products throughout the derivation. In order to treat the set $\{\hat{E}_n\}$ as a discrete (countable) basis, a convenient theoretical trick is to place the system in a box with absorbing boundary layers in which the absorption turns on more and more gradually. This procedure also gives the states $\hat{E}_n$ finite norms $(\hat{E}_n, \hat{E}_n)$. Because the operator is non-Hermitian, completeness of the basis can break down at an “exceptional point” (EP) [65,86], but EPs are not generically present; they must be forced by careful tuning of parameters. Therefore, we assume completeness in this paper and will treat the influence of EPs (self-orthogonal modes) as a limiting case in a future paper, as discussed in Sec. VIII.

Let $G(\omega, x, x')$ be the Green’s function of the operator $\mathcal{L}_\omega$, defined via $\mathcal{L}_\omega \hat{G}(\omega, x, x') = \delta(x - x')$ [90]. Given the complete set of eigenfunctions and eigenvalues $\{\hat{E}_n, \lambda_n\}$, the Green’s function can be expressed as the sum [90]

$$G(\omega, x, x') = \sum_n \frac{\hat{E}_n(x) \hat{E}_n^\dagger(x')}{\lambda_n(\omega)}.$$  

(A26)

Each lasing mode is associated with an eigenvalue $\lambda_n(\omega)$ of $\mathcal{L}_\omega$, which has a zero at a real frequency $\omega = \omega_n$. Consequently, $G(\omega, x, x')$ has a pole at $\omega_n$ and at frequencies near $\omega_n$, it is dominated by a single term in the sum. Expanding $\lambda_n(\omega)$
around the pole $\lambda_\mu(\omega) \approx (\omega - \omega_\mu)\lambda_\mu'$ (where $\lambda_\mu' \equiv \frac{\partial \lambda_\mu}{\partial \omega}|_{\omega_\mu}$), we obtain

$$G_{\mu\mu}(\omega, x, x') \approx \frac{E_{\mu}(x)E_{\mu}^T(x')}{(\omega - \omega_\mu)\lambda_\mu'} \int dx' E_{\mu}^2(x') = e^{-\omega_\mu t}. \quad (A27)$$

In order to evaluate $\lambda_\mu'$, let us rewrite $L_\omega$ as $L_\omega \approx L_0 + V(\omega)$, where $L_0 \equiv \nabla \times \nabla \times -\omega^2 \hat{\varepsilon}_0(\omega)$ and $V(\omega) \equiv -[\omega^2 \hat{\varepsilon}_0(\omega)]'_{\mu}(\omega - \omega_0)$. According to the Hellmann-Feynman theorem, the derivative of the eigenvalue $\lambda_\mu(\omega)$ with respect to $\omega$ is given by

$$\lambda_\mu' = \frac{\int dx E_{\mu}^2(x)[-\omega^2 \hat{\varepsilon}_0(\omega)]'_{\mu}}{(\omega - \omega_\mu)\int dx E_{\mu}^2(x)}, \quad (A28)$$

and substituting (A28) in (A27), we find that for frequencies near the resonances $\omega \approx \omega_\mu$, the Green’s function is approximately

$$G_{\mu\mu}(x, x', \omega) \approx \frac{E_{\mu}(x)E_{\mu}^T(x')}{(\omega - \omega_\mu)} \int dx' E_{\mu}^2(x') = e^{-\omega_\mu t}. \quad (A29)$$

5. The N-SALT TCMT equations

Having derived an expression for the Green’s function, the noise-driven field can be found by integrating the Green’s function over the source terms $F_{\text{NL}}(\omega, x, \omega)$ and $F_S(\omega, \omega)$:

$$\hat{E}_\mu(x, \omega) = \sum_v \omega_\mu^2 \int dx' G(x, x', \omega) \hat{\chi}_v(\omega_\mu, \omega_0) \hat{D}_v^* \hat{E}_\mu. \quad \text{and}$$

$$+ \int dx' G(x, x', \omega) \hat{F}_S. \quad (A30)$$

In the first term on the right-hand side, we approximate $\omega \approx \omega_\mu$ because the correction term is $O((\omega - \omega_\mu)\Delta \omega)$, which is second order in the noise. Substituting the single-pole approximation (A29) in (A30) yields

$$\hat{E}_\mu = \sum_v \omega_\mu^2 \frac{E_{\mu}(x, \omega_\mu, \omega_0)\hat{E}_{\mu}^2(x)}{(\omega - \omega_\mu) \int dx E_{\mu}^2(x)[\omega^2 \hat{\varepsilon}_0(\omega)]'_{\mu}} \hat{D}_v^* \hat{E}_\mu + \int \frac{dx'}{\omega - \omega_\mu} \hat{F}_S(x', \omega) E_{\mu}(x, \omega_\mu, \omega_0) \hat{E}_{\mu}^2(x'[\omega^2 \hat{\varepsilon}_0(\omega)]'_{\mu}). \quad (A31)$$

Finally, multiplying both side by $\omega - \omega_\mu$ and taking the inverse Fourier transform, we arrive at the N-SALT TCMT equations, which govern the evolution of the modal amplitudes $a_\mu$:

$$\dot{a}_\mu = \sum_v \int dx c_\mu(x) \gamma(x) \int dt' e^{i\omega x(x', t')} \times [a_\nu(t') - |a_\nu(t')|^2] a_\mu + f_{\mu}(t). \quad (A32)$$

The nonlinear coupling coefficient is

$$c_\mu(x) \equiv -i\omega_\mu^2 \frac{\chi_v(\omega_\mu, \omega_0) E_{\mu}^2(x)}{\int dx' E_{\mu}^2(x)[\omega^2 \hat{\varepsilon}_0(\omega)]'_{\mu}}. \quad (A33)$$

and the Langevin force is

$$f_{\mu}(t) \equiv \frac{i}{\mu} \int \frac{dx'\hat{F}_S(x', t') e^{-i\omega_\mu t}}{\int dx' E_{\mu}^2(x'[\omega^2 \hat{\varepsilon}_0(\omega)]'_{\mu}}. \quad (A34)$$

APPENDIX B: LINWIDTH OF THE MULTIMODE TIME-DELAYED MODEL

In this section, we calculate the laser linewidth for the multimode time-delayed model by generalizing the solution strategy of Sec. V in the text. We begin our analysis with the discretized time-delayed N-SALT TCMT equation (35) (repeated here for convenience):

$$\dot{a}_\mu = \sum_v C_{\mu v} \int dt' e^{-\gamma(t-t')} |a_{\nu}(t')|^2 - |a_{\nu}(t')|^2 |a_\mu + f_{\mu} |. \quad (B1)$$

Following the approach of Sec. V A, we linearize (B1) by expanding the mode amplitudes $a_\mu$ around their steady-state values, $a_\mu = (a_{\mu 0} + \delta a_\mu)e^{i\phi_\mu}$ (where $\delta a_\mu \ll a_{\mu 0}$), and we omit the terms $O(\delta a_\mu^2)$, Then, we introduce additional variables $\xi_{\mu v}^k$, where

$$\xi_{\mu v}^k = \gamma_k \int dt' e^{-\gamma_k(t-t')} \delta a_{\mu v}(t'), \quad (B2)$$

where $\xi_{\mu v}^k$ is the time-averaged amplitude deviation of mode $\mu = 1 \ldots M$ from steady state at the spatial point $k = 1 \ldots N$. Having introduced the auxiliary variables $\xi_{\mu v}^k$, the set of integro-differential equations (B1) turns into a linear system of ODEs, which we solve by applying several linear-algebraic transformations to obtain a compact expression for the covariance matrix, as described in detail below.

Introducing the vector $\Phi_{\mu} \equiv a_{\mu 0} \Phi_{\mu}$, the linear system of ODEs is conveniently written as

$$\dot{\delta}_\mu = -\sum_{v k} (2a_{\mu 0} a_{\nu 0} \text{Re}[C_{\mu v}^k]) \xi_{\mu v}^k + f_{\mu}^R, \quad (B3)$$

$$\dot{\Phi}_{\mu} = -\sum_{v k} (2a_{\mu 0} a_{\nu 0} \text{Im}[C_{\mu v}^k]) \xi_{\mu v}^k + f_{\mu}^I, \quad (B4)$$

To simplify the notation further, we introduce the $M \times M$ matrices $A^k$ and $B^k (k = 1 \ldots N)$ with entries

$$A^k_{\mu v} = 2a_{\mu 0} a_{\nu 0} \text{Re}[C_{\mu v}^k], \quad (B6)$$

$$B^k_{\mu v} = 2a_{\mu 0} a_{\nu 0} \text{Im}[C_{\mu v}^k]. \quad (B7)$$

and we rearrange the set of equations Eqs. (B3) and (B5) in a matrix form [compare with (24,25)]:

$$d \delta \mu \over dt = -\sum_k A_k^k \xi^k + f^R, \quad (B8)$$

$$d \Phi \mu \over dt = -\sum_k B_k^k \xi^k + f^I, \quad (B9)$$

The autocorrelation matrix of the phase vector $\Phi$, which we calculate in this section, is determined by the autocorrelation matrix of the Langevin force,

$$\langle f(t) f^T(t') \rangle = \Re \delta(t - t'). \quad (B11)$$

063806-17
In order to compute $\langle \Phi \Phi^T \rangle$, we solve (B9) by straightforward integration. We find that the phase covariance matrix is a sum of a “pure” phase-diffusion term, proportional to $\frac{R}{2}$, and an amplitude-phase coupling term, proportional to $J$,

$$\langle \Phi(t)\Phi^T(0) \rangle = \left( \frac{R}{2} + J \right)|t|,$$

where we have introduced the shorthand notation

$$J = \frac{1}{|t|} \sum_{kl} B_k^T \int dt' dt'' \langle \xi_k(t')\xi_l(t'') \rangle B_l^T$$

(B12)

for the second term, which is responsible for the generalized $\alpha$ factor.

In the remainder of this section, we calculate $J$. First, we solve the set of ODEs for $\xi^k$ and $\delta$ (B8) and (B10), and then we substitute the solution for $\delta$ into (B13) and evaluate the integrals. To this end, we begin by rewriting the equations for $\xi^k$ and $\delta$ more compactly. We define the $[(N + 1)M] \times 1$ vectors $x$ and $F$ and the $[(N + 1)M] \times [(N + 1)M]$ matrix $K$:

$$x(t) = \left( \begin{array}{c} \xi^1(t) \\ \vdots \\ \xi^N(t) \end{array} \right), \quad F(t) = \left( \begin{array}{c} F^1(t) \\ \vdots \\ F^N(t) \end{array} \right).$$

$$K = \left( \begin{array}{ccc} 0 & A_1 & A_2 & \cdots & A_N \\ -A_1 & 0 & 0 & \cdots & 0 \\ A_2 & -A_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_N & \cdots & 0 & \cdots & 0 -A_{N-1} \end{array} \right)$$

where $A_k$ are block-diagonal $M \times M$ matrices with $\gamma_k$ on the diagonal entries and the zeros in the definition of $K$ are block $M \times M$ zero matrices. Using these definitions, the equations for $\xi^k$ and $\delta$ Eqs. (B8) and (B10) can be written as

$$\frac{dx}{dt} = -Kx + F.$$  

The solution of (B15) is

$$x_m(t) = \int_0^t dt' \sum_{\rho} \int_0^t dt'' \langle e^{-K(t - t'')} \rangle_{\rho} F_{\rho}(t').$$

(B16)

and, in particular, the solution for $\xi^k_m$ is

$$\xi^k_m(t) = \int_0^t dt' \sum_{s=1}^M \int_0^t dt'' \langle e^{-K(t - t'')} \rangle_{Ms,s} f_s \langle \xi^k(t') \rangle.$$  

(B17)

For ease of notation, let us denote the $(k + 1)\text{th}$ $M \times M$ block in the first column of the matrix $e^{-K(t - t'')}$ by the shorthand notation $[e^{-K(t - t'')}]_{k+1,1}$, so that $\xi^k = \int_0^t dt' \langle e^{-K(t - t'')} \rangle_{k+1,1} F(t')$. Substituting the expression for $\xi^k$ into $J$ and using the autocorrelation function of the Langevin force (B11), we obtain

$$J = \frac{1}{|t|} \sum_{kk'} B_k^T \int dt' dt'' ds' \langle e^{-K(t' - s')} \rangle_{k+1,1} \int dt'' ds' \langle e^{-K(t - s')} \rangle_{k+1,1} B_l^T.$$  

(B18)

We proceed (not shown) by diagonalizing the matrix $K$ and evaluating the integrals in (B18). (The intermediate steps depend on the eigenvalues of $K$ and the matrix of eigenvectors, but the final result can be expressed in terms of the matrix inverse $K^{-1}$). In the long-time limit, we keep the leading-order term (which grows linearly in time) and we obtain

$$J = \left( \sum_k B_k \sum_{k+1} \langle \xi^k_{k+1,1} \rangle \right) \frac{R}{2} \left( \sum_k \langle \xi^k_{1,1} \rangle \right)^T.$$  

(B19)

In order to complete the derivation of the linewidth formula, we use the identity

$$[K]^{-1}_1 = \left( \sum_j A_j \right)^{-1},$$  

(B20)

which we prove below. Noting that $A_{\mu \nu} = \sum_k A_{\mu \nu}^k$, and $B_{\mu \nu} = \sum_k B_{\mu \nu}^k$, and using the identity (B20), we find that (B19) reduces to

$$J = BA^{-1} \frac{R}{2} (BA^{-1})^T,$$

which completes the derivation of the linewidth formula in the most general time-delayed model. In particular, and somewhat remarkably, the $\gamma$ terms completely cancel in the computation of the first column of the matrix inverse and drop out of the final result.

Proof of the identity (B20). We use the Schur complement [91] for the lower-left corner of a matrix inverse:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A - BD^{-1}C \end{pmatrix}^{-1}.$$  

(A and $D$ need to be square matrices). Decomposing the matrix $K$ into the blocks

$$A = (0), \quad B = (A_1, A_2, \cdots, A_N)$$

$$C = \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_N \end{pmatrix}, \quad D = \begin{pmatrix} -\Delta_1 & 0 & \cdots & 0 \\ 0 & -\Delta_2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & -\Delta_N \end{pmatrix},$$

we can calculate the lower-left corner of $[K]^{-1}_1$:

$$-D^{-1}C(A - BD^{-1}C)^{-1} = (D^{-1}C)(B(D^{-1}C))^{-1}$$

$$= -\begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_N \end{pmatrix} \begin{pmatrix} (A_1, A_2, \cdots, A_N) \end{pmatrix}^{-1}$$

$$= -\begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_N \end{pmatrix} \begin{pmatrix} \sum \Delta_1 \end{pmatrix}^{-1}.$$  

Therefore, we obtain $[K]^{-1}_1 = (\sum_j A_j)^{-1}$.  

063806-18
APPENDIX C: LEMMA FROM SEC. VLC—\([B \Lambda^{-1}]_{ij} \propto \frac{2\pi}{\ell_{ij}}\]

In Sec. VI.B, we present a multimode ST relation, which states that the linewidths are proportional to a sum of inverse output powers of all the other modes. This result arises from a lemma which we prove here. We use the standard matrix-inverse formula [92],

\[ \Lambda^{-1} = \frac{1}{\text{det} \Lambda} \text{adj} \Lambda, \]  

(C1)

where the adjugate matrix is defined as

\[ \text{adj} \Lambda = [(-1)^{i+j} M_{ij}]^T. \]  

(C2)

\( \Lambda \) is the cofactor matrix, i.e., the matrix whose \((i, j)\) entry is the determinant of the \((i, j)\) minor of \( \Lambda \) (which is the matrix obtained from \( \Lambda \) by deleting the \(i\)th row and the \(j\)th column). From the definition of \( \Lambda \) (i.e., \( \Lambda_{ij} \equiv \text{Re}[C_{ij} \alpha_{0j} \alpha_{0i}] \)), it follows that

\[ M_{ij} = \left( \prod_{k \neq i,j} a_{0k}^2 \right) a_{0i} a_{0j} Q_{ij}, \]  

(C3)

where \( Q_{ij} \) (and later \( Q \)) denote constants that may depend on \( i \) and \( j \), but are independent of the modal amplitudes. Note also that

\[ \text{det} \Lambda = \prod_k a_{0k}^2 Q. \]  

(C4)

Using (C1)–(C4), we obtain

\[ \Lambda^{-1}_{ij} = \frac{1}{a_{0i} a_{0j}} Q_{ij}. \]  

(C5)

Therefore, one can easily see that the lemma follows, since

\[ [B \Lambda^{-1}]_{ij} \propto \sum_k a_{0i} a_{0k} \frac{1}{\epsilon_{\alpha_{0k}}} \propto \frac{a_{0i}}{\epsilon_{\alpha_{0i}}}. \]

APPENDIX D: COMPARISON WITH THE SCATTERING-MATRIX LINEWIDTH FORMULA

In a recent scattering-matrix-based linewidth theory [14], Pillay et al. obtain a formula for the linewidth of a 1D laser system, expressed in terms of integrals over the modes which solve the nonlinear SALT equations. In this Appendix, we prove that their formula (which applies to 1D systems) is equivalent to our linewidth formula (3) (except that their formula gives a spatially averaged incomplete-inversion factor and omits the \( \alpha \) factor).

In the scattering-matrix approach, the lasing modes are described as purely outgoing wave functions \( \psi_0 \), which satisfy the nonlinear SALT equation

\[ \nabla \times \nabla \times \psi_0(x) - \omega_0^2 \varepsilon(x, \omega_0) \psi_0(x) = 0, \]  

(D1)

and can be expressed as a superposition of outgoing channel modes \( u_{\mu} \) outside of the laser region

\[ \psi_0(x) = \sum_k b_k u_k(x, \omega_0) \quad \text{for} \quad x \notin C. \]  

(D2)

\( C \) denotes the scattering region (i.e., \( \varepsilon = 1 \) for \( x \notin C \)). [Note that \( \psi_0(x) \) is precisely the same the mode \( E_0(x) \) (which was used in Sec. II) inside the cavity region.] The outgoing mode amplitudes \( b \) are normalized to the value of \( \psi_0 \) at the cavity boundary \((x = L)\),

\[ b^T b = \psi_0^2(L). \]  

(D3)

The apparent difference between our formula and the linewidth formula in [14] is that the integral term in the denominator of our linewidth formula (3) is replaced with a sum of two terms in the scattering-matrix approach,

\[ \int_\text{space} dx \left[ \varepsilon \omega_0 + \frac{\varepsilon_0^2}{2} \frac{d\varepsilon}{d\omega_0} \right] \psi_0^2 \]  

\[ \longrightarrow \frac{i b^T b}{2} + \int_\text{dx} \left[ \varepsilon \omega_0 + \frac{\varepsilon_0^2}{2} \frac{d\varepsilon}{d\omega_0} \right] \psi_0^2. \]  

(D4)

In order for the two formulas to agree, we need to show that

\[ \omega_0 \int_L^\infty dx \psi_0^2(x) = \frac{i b^T b}{2} \]  

(D5)

(where we have used the fact that \( \varepsilon = 1 \) outside the cavity region). We show that the latter condition (D5) holds for any solution \( \psi_0 \) of (D1), which satisfies outgoing boundary conditions. One way to impose outgoing boundary conditions is to invoke the limiting-absorption principle (i.e., add loss to eliminate incoming waves from infinity and take the limit of infinitesimal absorption at the end of the calculation [93,94]). Formally, we define the integral on the left-hand side of (D5) as

\[ \int_L^\infty dx \psi_0^2(x) \equiv \lim_{s \to 0^+} \int_L^\infty dx e^{-sx} \psi_0^2(x). \]  

(D6)

By substituting \( \psi_0(x) = e^{ikx} \) into (D6) and taking the limit of \( s \to 0^+ \), we obtain \( \int_L^\infty dx \psi_0^2(x) = \frac{1}{2\pi} e^{ikL} = \frac{i}{2\pi} b^T b \), and since \( \omega_0 = ck_0 \) this finishes the proof of (D3) (with the units convention of \( c = 1 \)).

APPENDIX E: ZERO-POINT FLUCTUATION CANCELLATION

The hyperbolic cotangent factor in the FDT (19) arises as a sum of a Bose-Einstein distribution and a \( k \) factor stemming from quantum zero-point (ZP) fluctuations [70,71], and this is why it does not vanish in the limit of zero temperature (\( \beta \to \infty \)). However, it turns out that contribution of this ZP term cancels in the linewidth formula, as was shown by Henry and Kazarinov [72] from a quantum-operator viewpoint, and it is convenient to explicitly subtract the ZP term from the hyperbolic cotangent as in (3) and (21). Here we provide a purely classical explanation for why this cancellation occurs and why it is important to perform the explicit subtraction in order to eliminate a subtlety arising from the definition of outgoing boundary conditions.

The FDT has a hyperbolic cotangent factor, and when we apply the FDT to find the \( \langle f_x f^*_x \rangle \) correlation function in Sec IV, the same hyperbolic cotangent factor arises in the \( R \) integral,
appearing in the form

\[
\int dx |E_\mu|^2 \text{Im} \varepsilon_\mu(\omega_\mu) \left( \frac{\hbar \omega_\mu \beta}{2} \right)
\]

\[
= \int dx |E_\mu|^2 \text{Im} \varepsilon_\mu(\omega_\mu) \left[ \left( \frac{1}{2} \coth \left( \frac{\hbar \omega_\mu \beta}{2} \right) - \frac{1}{2} \right) + \frac{1}{2} \right]
\]

(E1)

for a lasing mode \(\mu\), where we have trivially added and subtracted the ZP \(1/2\) factor from \(\coth\). Now we wish to analyze the final \(1/2\) term, which is the integral \(\int dx |E_\mu|^2 \text{Im} \varepsilon_\mu(\omega_\mu)\).

Before we treat outgoing boundary conditions, let us consider the simpler case of a laser surrounded by an explicit absorbing medium, as in [72]. (This is also the situation in more recent computational models, for which one uses a finite spatial domain surrounded by absorbing layers [21].) For any steady-state lasing mode (real \(\omega_\mu\)), the net gain + loss is zero, but \(\frac{1}{2} \int dx |E_\mu|^2 \text{Im} \varepsilon_\mu(\omega_\mu)\) is proportional to the net power absorbed or gained by the electric field [95] and hence this integral is zero. Therefore, in such a case, whether or not we include the \(1/2\) factor, is irrelevant, because the \(\pm 1/2\) terms integrate to zero.

However, a subtlety arises in this integral in the common case where the laser is surrounded by an infinite zero-temperature \((\beta = \infty)\) lossless medium with outgoing radiation boundary conditions. Outgoing boundary conditions can be defined mathematically by the limiting absorption principle [93,94]: One takes the lossless medium to be the limit of a lossy medium as the losses go to zero from above, which can be expressed by writing \(\varepsilon\) as \(\varepsilon + i 0^+\). Just as in Appendix D, the correct approach is to take the lossless limit after solving the problem, i.e., the \(0^+\) limit is taken outside of the integral. Before we take this limit, it makes no difference whether the \(1/2\) factor is included, just as above: It integrates to zero. However, after we take the lossless limit, there is no explicit absorbing region \((\Im \varepsilon > 0)\) in the integral (the absorption has been “moved to infinity” in some sense), so if we perform the coth integral without subtracting \(1/2\) then we would obtain an incorrect contribution from the ZP fluctuations in the gain medium (which should have been canceled). Instead, if we integrate against \(\pm \coth - 1/2\), the result is correct without requiring any explicit contribution from the absorbing boundary conditions.

Note that if the laser is surrounded by an infinite lossless medium at a positive temperature, then there is a nonzero contribution of incoming thermal radiation to the linewidth [2,3]. This can be included in one of two ways. In practice, we typically solve the SALT equations in a finite computational box with an explicit absorbing region, in which case no modification to our linewidth formula is required: One simply assigns the ambient temperature to the absorbing region. If, on the other hand, the outgoing boundary conditions are imposed in some other way (e.g., semianalytically as in earlier SALT work [19,20,43,96]), then an explicit source term must be added to account for incoming thermal radiation, as in previous works [97].


[78] See Eq. (45) in Ref. [29].


[80] See Eq. (A4) in Ref. [20].


