# Single-mode instability in standing-wave lasers: The quantum cascade laser as a self-pumped parametric oscillator

Tobias S. Mansuripur,[1] Camille Vernet,[2,3] Paul Chevalier,[2] Guillaume Aoust,[2,4] Benedikt Schwarz,[2,5] Feng Xie,[6] Catherine Caneau,[7] Kevin Lascola,[6] Chung-en Zah,[6] David P. Caffey,[8] Timothy Day,[8] Leo J. Missaggia,[9] Michael K. Connors,[9] Christine A. Wang,[9] Alexey Belyanin,[10] and Federico Capasso[2,*]

[1]*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*
[2]*John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*
[3]*Ecole Polytechnique, 91123 Palaiseau, France*
[4]*ONERA, The French Aerospace Lab, 91123 Palaiseau, France*
[5]*Institute of Solid State Electronics, TU Wien, 1040 Vienna, Austria*
[6]*Thorlabs Quantum Electronics (TQE), Jessup, Maryland 20794, USA*
[7]*Corning, Inc., Corning, New York 14831, USA*
[8]*Daylight Solutions, Inc., San Diego, California 92128, USA*
[9]*Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, Massachusetts 02420, USA*
[10]*Department of Physics and Astronomy, Texas A & M University, College Station, Texas 77843, USA*

We report the observation of a clear single-mode instability threshold in continuous-wave Fabry-Perot quantum cascade lasers (QCLs). The instability is characterized by the appearance of sidebands separated by tens of free spectral ranges (FSR) from the first lasing mode, at a pump current not much higher than the lasing threshold. As the current is increased, higher-order sidebands appear that preserve the initial spacing, and the spectra are suggestive of harmonically phase-locked waveforms. We present a theory of the instability that applies to all homogeneously broadened standing-wave lasers. The low instability threshold and the large sideband spacing can be explained by the combination of an unclamped, incoherent Lorentzian gain due to the population grating, and a coherent parametric gain caused by temporal population pulsations that changes the spectral gain line shape. The parametric term suppresses the gain of sidebands whose separation is much smaller than the reciprocal gain recovery time, while enhancing the gain of more distant sidebands. The large gain recovery frequency of the QCL compared to the FSR is essential to observe this parametric effect, which is responsible for the multiple-FSR sideband separation. We predict that by tuning the strength of the incoherent gain contribution, for example by engineering the modal overlap factors and the carrier diffusion, both amplitude-modulated (AM) or frequency-modulated emission can be achieved from QCLs. We provide initial evidence of an AM waveform emitted by a QCL with highly asymmetric facet reflectivities, thereby opening a promising route to ultrashort pulse generation in the mid-infrared. Together, the experiments and theory clarify a deep connection between parametric oscillation in optically pumped microresonators and the single-mode instability of lasers, tying together literature from the last 60 years.

## I. INTRODUCTION

In the last decade, significant efforts have spurred the understanding of high-$Q$ optically pumped microresonators. A monochromatic external pump beam is coupled to a mode of the microresonator, and at sufficient pump power the third-order $\chi^{(3)}$ Kerr nonlinearity, responsible for the intensity-dependent refractive index, couples the pumped mode to fluctuations at other frequencies, which leads to interesting physics. Starting from an initial demonstration of third-order optical parametric oscillation (OPO) [1,2], in which the pump beam provides sufficient parametric gain to allow a few pairs of sidebands to oscillate, this technique has been extended to generate wide-spanning frequency combs [3,4], and most recently temporal solitons [5,6]. The many degrees of freedom one can manipulate in these systems, such as the group velocity dispersion (GVD), the free spectral range of the resonator, the detuning of the pump frequency relative to the cold-cavity

mode that it pumps, and the pump power, among others, have provided a rich nonlinear optical playground to observe diverse physical phenomena.

A laser, much like an OPO, is an optical resonator in which circulating monochromatic light reaches high intensity, the difference being that the light is internally generated rather than externally injected. Furthermore, the very gain medium that allows for lasing, simultaneously provides a third-order nonlinearity, the population pulsation (PP) nonlinearity [7]. The PP nonlinearity is an intrinsic property of any two-level system that interacts with near-resonant amplitude-modulated (AM) light: the radiative transition rate between the states, and therefore the population of each state, is temporally modulated by the AM light, resulting in so-called population pulsations that act back on the light field in a nonlinear way. The laser therefore contains the two ingredients, high-intensity light and a nonlinearity, necessary for parametric oscillation. Indeed, in the late 1960s the importance of PPs in determining the above-threshold spectral evolution of a homogeneously broadened, traveling-wave laser was realized. At the laser threshold, one mode, which we call the primary mode, begins

*capasso@seas.harvard.edu

to lase and as the current is increased the population inversion remains clamped to its threshold value. It was first thought that this clamping should prevent any other mode from reaching the oscillation threshold. This reasoning, however, neglects the fact that when a photon of a different frequency is spontaneously emitted in the presence of the primary lasing field, a beat note, i.e., an intensity modulation at the difference frequency of the two fields, is created. The beat note creates a PP that provides a parametric contribution to the gain of the spontaneously emitted photon. At a sufficiently high pumping level known as the instability threshold, this parametric gain can, despite the fact that the population inversion is clamped to its threshold value, allow two sidebands to overcome the loss. The separation of these sidebands from the primary mode is related to the Rabi frequency induced by the primary mode. This effect is responsible for both the Haken-Risken-Schmid-Weidlich (HRSW) instability [8,9] and the Risken-Nummedal-Graham-Haken (RNGH) instability [10,11]. Many years later, insightful work properly identified the fundamental role of PPs in the single-mode laser instabilities [12–18] and also chaos [19]. [We note that PPs are important not only for inverted media. Historically, their effects were first appreciated in microwave spectroscopy pump-probe experiments by Autler and Townes [20] in 1955, and soon came to be known as the ac Stark effect. Through the late 1960s and 1970s, significant work on sideband amplification [21], resonance fluorescence [22,23], and the Mollow scattering triplet [24–28] culminated in the "dressed" description of atoms in strong fields [29]. In the 1980s, the PP nonlinearity was cast in the language of nonlinear optics and applications such as four-wave mixing [30], phase conjugation [31], and optical bistability [14] were explored.]

Both the HRSW and RNGH single-mode instabilities apply to homogeneously broadened traveling-wave lasers, and predict the appearance of sidebands on the primary lasing mode, as shown in Fig. 1(a). We remark that, in general, the temporal behavior of an electric field that contains three equally spaced frequencies can be more amplitude modulated or frequency modulated (FM), depending on the spectral phase, as shown in Fig. 1(b). One can think of the intensity modulation (in other words, the beat note) of the AM and FM fields as resulting from the sum of two phasors rotating at frequency $\delta\omega$, each of which is created by the beat between a sideband and the primary mode. As shown in Fig. 1(c), the two phasors either constructively interfere to create a large intensity modulation (AM) or destructively interfere to eliminate the intensity modulation (FM). In both the HRSW and RNGH instabilities, the three-wave field is by necessity AM; a constant-intensity FM field would not create the PP and the resulting parametric gain that is required by the sidebands to reach the lasing threshold. In the HRSW case, which applies to low quality-factor cavities for which the photon lifetime is shorter than the atomic decay time, the sideband separation is smaller than the mode spacing, or free spectral range (FSR), of the cavity. All three lasing frequencies fall within a single cold-cavity resonance, which is made possible by a region of anomalous dispersion created by the PP [12]. In the RNGH instability, which applies to higher quality-factor cavities, the sidebands must coincide with cold-cavity modes in order to satisfy the round-trip phase condition, resulting in a separation
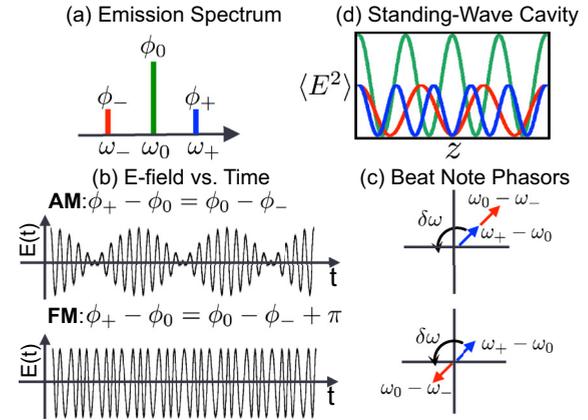


FIG. 1. (a) The emission spectrum at the instability threshold comprises a primary mode and two weak sidebands. (b) The temporal behavior of the field $E(t)$ depends on the relative phases of the three modes $\phi_-$, $\phi_0$, and $\phi_+$, and shown are the AM and FM configurations. (c) The AM and FM fields can be understood in terms of the constructive and destructive addition of two beat note phasors, where each phasor represents a contribution to the intensity modulation at the difference frequency $\delta\omega$ resulting from the superposition of each sideband with the primary mode. (d) In a standing-wave cavity, the intensity of each mode varies with position, and the spatial modes corresponding to different frequencies do not perfectly overlap.

that is an integer multiple of the FSR. An important corollary of this requirement is that to observe the effect of the PPs, the FSR must be smaller than the gain recovery frequency (i.e., inverse of the gain recovery time $T_1$). Why? The gain recovery time determines the shortest time scale at which the population inversion can respond to an intensity modulation; therefore, the amplitude of the PP is only significant for sidebands detuned by an amount close to or smaller than $1/T_1$. If the FSR is greater than $1/T_1$, then all cavity modes are too far from the primary mode to generate a PP with an amplitude large enough to make the sidebands unstable. Ideally, the FSR should be significantly smaller than $1/T_1$ so that the FP modes densely populate the parametric gain lobe, increasing the probability of satisfying the instability condition. Provided this condition is met, the RNGH instability predicts that a traveling-wave laser with rapid dephasing must be pumped nine times above threshold before the instability appears. Experimental observations of a rhodamine dye ring laser [16] showed signatures of an RNGH-type instability, with two key differences: the instability threshold was only fractionally higher, not nine times higher, than the lasing threshold, and the sideband creation was accompanied by the disappearance of the primary mode. Efforts to explain the discrepancies between theory and experiment are well summarized in [17,32], but to our knowledge the discrepancy was never fully resolved.

In this work, we will investigate the single-mode instability in a standing-wave laser, shown schematically in Fig. 1(d). The distinguishing feature of the standing-wave laser is that the primary mode induces a population grating (PG) (as long as carrier diffusion is limited), an effect known as spatial hole burning (SHB). The gain of other cavity modes is no longer clamped above threshold, but continues to increase with the pumping. Therefore, the instability threshold can

be reached *without* the need for PP parametric gain. We call this an incoherent instability, and it occurs in media whose gain recovery time is too slow for PPs to occur (FSR $> 1/T_1$), such as diode lasers. In gain media with a fast recovery time (FSR $< 1/T_1$), both the incoherent gain and the parametric PP contribution to the gain must be considered. We will show that the PP parametrically suppresses the gain of nearby sidebands because low-frequency sidebands cause the population inversion to oscillate perfectly out of phase with the intensity modulation. On the other hand, the PP enhances the gain of larger-detuning sidebands, as occurs in the RNGH instability. Depending on the relative contributions of the incoherent and coherent gain, we show that the laser will either emit an FM or an AM waveform at the instability threshold, to either minimize or maximize the amplitude of the PPs. If the incoherent gain is large, nearby sidebands are favored and will yield FM emission to minimize the amount of parametric suppression. If the incoherent gain is small, larger-detuning sidebands are favored and will yield AM emission to maximize the amount of parametric enhancement. The possibility of both FM and AM emission from a standing-wave laser is a novelty not shared by the traveling-wave laser, which, as mentioned before, can only produce an AM waveform.

The quantum cascade laser (QCL) is precisely the kind of laser for which both the PG and PPs are important. An electron injected into the upper state has only a short picosecond lifetime during which to diffuse before it scatters to the ground state, not enough time to traverse the half-wavelength mid-infrared ($\lambda \sim 3$–$12 \mu$m) standing wave from node to antinode. Therefore, the PG is not washed out. Also, the FSR (typically 8 to 16 GHz) is much less than the gain recovery frequency ($1/T_1 \approx 1$ THz), so the population inversion has no difficulty following the beat notes in field intensity created when multiple modes lase simultaneously, yielding PPs. We report the discovery that continuous-wave (cw) Fabry-Perot (FP) QCLs reach a well-defined instability threshold, characterized by the appearance of sidebands whose separation from the primary mode can be several multiples of the cavity FSR. This mode skipping is a clear signature of the parametric PP interaction between the primary mode and the sidebands, which strongly suppresses sidebands at separations much smaller than the large gain recovery frequency of the QCL. The behavior is observed in QCLs that emit at wavelength 3.8, 4.6, and 9.8 $\mu$m, indicating that it is a universal feature of mid-infrared QCLs, independent of the specific band structure of the active region. The strength of the PG can be tuned by coating the facets to adjust their reflectivities. By comparing the measurements with the theory, we argue that QCLs with uncoated facets emit an FM waveform. A QCL with one high-reflectivity facet and a sufficiently low reflectivity of the other facet should in principle emit an AM waveform, and we provide preliminary evidence that this is indeed the case, demonstrating a QCL whose sidebands are separated from the primary mode by 46 FSR. While the PG and PP have been known to be important in QCLs, in previous work their effects were treated separately [33]. Instead, we emphasize that one should think of the PG, a spatial modulation of the inversion, and the PP, a temporal modulation of the inversion, as working *in tandem* to create a phase-locked multimode state at low pump power.

As the current is increased past the instability threshold, higher-order sidebands that preserve the initial spacing appear. This suggests that the FP-QCL can emit a harmonically phase-locked waveform without the need for any external modulation or additional nonlinear elements. Why have such spectra not been observed before, except in a few cases [34,35]? We have found the harmonic states to be extremely sensitive to optical feedback. Simply placing a collimating lens between the QCL and the spectrometer, even a poorly aligned, tilted lens with a focal length of a few cm, makes it difficult to observe the harmonic state, and instead yields the more familiar QCL spectrum in which all adjacent FP modes lase. It is also important to slowly increase the current, which allows for a smooth transition from the single mode to the harmonic regime. We argue that the harmonic state is an intrinsic regime of all QCLs. The fact that it has only been observed 15 years after the invention of the cw QCL is a testament to the destabilizing influence of optical feedback [36].

In the last few years, comb generation in a QCL on adjacent FP modes has been demonstrated [37–39], and the importance of parametric mode coupling is known [40,41]. (These devices all had multistage inhomogeneously broadened active regions, which distinguishes them from the devices in our work.) Because these combs have so far always comprised adjacent cavity modes (with the exception of [42]), consideration has only been given to the case where the fundamental frequency of the PPs equals the FSR. This low PP frequency strongly favors the emission of an FM waveform. The remarkable degree of freedom to skip modes, never before considered, means that the temporal periodicity of the PPs is no longer pinned to the cavity round-trip time (typically 60 to 120 ps), but is shortened by a factor equal to the number of modes skipped, which reaches 46 in one of our QCLs. This reduction of the period down to the order of the gain recovery time is the crucial feature that allows for the possibility of AM emission.

Finally, we emphasize the deep connection between the single-mode laser instability and mode proliferation in optically pumped microresonators. Both are cases of parametric oscillation that are initiated by a nonlinearity, either PP or Kerr, transferring energy from a pump beam to two sidebands. For a passive microresonator, the pump beam must be injected, while in the laser the pump beam is internally generated. This analogy, which we only begin to uncover here, can help guide future work toward understanding the rich emission spectra of QCLs. More broadly, both QCLs and microresonators exhibit the widespread phenomenon of modulation instability [43]. We hope that the advancement of the QCL can parallel the rapid progress seen in microresonators in the last decade, leading to a compact source of mid-infrared frequency combs for spectroscopy of trace gases and short pulse generation [44].

In Sec. II, we present the experimental results, which help to motivate the theory presented in Sec. III. In Sec. IV, we compare the theory with the measurements, and finally conclude in Sec. V.

## II. EXPERIMENT

All four devices used in this study are cw, buried heterostructure, FP-QCLs. Our device naming convention

TABLE I. Summary of relevant parameters of the devices used in this study. $n_{\text{eff}}$, $T_2$, $\delta\omega_{\text{sb}}$, and $J_{\text{sb}}$ were measured quantities. $T_{\text{up}}$ and $d$ were calculated from the band structure for TL-4.6 and DS-3.8, and taken from [46] for LL-9.8, and $\gamma_D$ was calculated assuming $D = 77$ cm$^2$/s [49].

| Device | $n_{\text{eff}}$ | $d$ (nm · $e$) | $T_{\text{up}}$ (ps) | $\gamma_D$ | $T_2$ (fs) | $\delta\omega_{\text{FSR}}$ (GHz) | $\delta\omega_{\text{sb}}$ (GHz) | $J_{\text{sb}}/J_{\text{th}}$ |
|---|---|---|---|---|---|---|---|---|
| LL-9.8 | 3.43 | 3 | 0.54 | 0.93 | 81 | 92 | 642 | 1.14 |
| TL-4.6 | 3.23 | 1.63 | 1.7 | 0.49 | 74 | 48 | 1259 | 1.17 |
| TL-4.6:HR/AR | 3.25 | 1.63 | 1.7 | 0.49 | 74 | 48 | 2216 | 1.22 |
| DS-3.8 | 3.25 | 1.5 | 1.74 | 0.40 | 43 | 49 | 977 | 1.12 |

identifies the provider of the device (LL: MIT Lincoln Laboratory, TL: Thorlabs, DS: Daylight Solutions) followed by the emission wavelength in microns. The active region of device LL-9.8 is a double phonon resonance design using lattice-matched $Ga_{0.47}In_{0.53}As/Al_{0.48}In_{0.52}As$, grown by metal-organic chemical vapor deposition, with the well-known layer structure of [45] (with a nominal doping of $n = 2.5 \times 10^{18}$ cm$^{-3}$), for which extensive band-structure calculations have been done [46]. The device length is 3 mm and width is 8 $\mu$m. Devices TL-4.6, TL-4.6:HR/AR, and DS-3.8 were grown using strained $Ga_xIn_{1-x}As/Al_yIn_{1-y}As$ and are described in [47], although the layer sequence is not given. The length is 6 mm and width is 5 $\mu$m for these three devices. Both facets are left uncoated for LL-9.8, TL-4.6, and DS-3.8. The only coated device is TL-4.6:HR/AR, which has a high-reflectivity (HR) coating on the back facet ($R \approx 1$) and an antireflection (AR) coating on the front facet ($R \approx 0.01$), but is otherwise nominally identical to TL-4.6. Far-field measurements indicate that all devices exhibit single lateral-mode emission over the full range of applied current. It is worth mentioning that the short-wave QCLs, DS-3.8 and TL-4.6, have positive GVD and the long-wave device LL-9.8 has negative GVD. We expect this because their wavelengths lie on opposite sides of the zero-GVD point of InP, but we have also confirmed this in Appendix B using the subthreshold measurement method of [48]. Some relevant parameters for each device are given in Table I: the effective refractive index $n_{\text{eff}}$ is determined from the FP-mode spacing of the measured spectra; the dipole moment $d$ and the upper-state lifetime $T_{\text{up}}$ are calculated from the band structure; the dephasing time $T_2$ is determined from a Lorentzian fit to either an electroluminescence or far-subthreshold measurement of each device. The output power of each laser was measured with a calibrated thermopile (Ophir 3A-QUAD) placed close to the facet; the total output power is plotted in Fig. 2, which is obtained from the front facet only for TL-4.6:HR/AR and by doubling the single-facet power of the uncoated lasers.

Our goal was to precisely examine the spectral evolution of the QCL with increasing current, from the single-mode to the multimode regime. Specifically, we wanted to answer the following question: At what pumping level does a second mode start to lase, and what is the relationship between the second frequency and the first? To answer this question, we would begin each measurement with the laser driven at a current beneath the laser threshold. The current was then slowly increased in steps of 1 mA (at a rate of roughly 2 mA per second), and the spectrum was monitored using a Fourier transform infrared (FTIR) spectrometer (Bruker Vertex 80v), with either an InSb detector (for DS-3.8 and TL-4.6) or HgCdTe detector (for LL-9.8), both cryogenically cooled.

The current was supplied by a low-noise driver (Wavelength Electronics QCL1500 or QCL2000), and the temperature of the copper block beneath the QCL was stabilized to 15 °C. The slow rate of increase of the current was necessary to precisely identify the instability threshold, and also to prevent rapid temperature variations. To completely eliminate the possibility of optical feedback due to reflections from optical elements outside the laser cavity, the QCL was placed about 40 cm from the entrance window to the FTIR and its output was not collimated with a lens, but simply allowed to diverge. The high power of the devices and the sensitivity of the detectors was sufficient to measure spectra despite the small fraction of collected optical power.

Spectra measured in this manner are shown in Fig. 3 for the three uncoated devices. Each spectrum is normalized to its own maximum and plotted on a logarithmic scale covering 40 dB of intensity variation. All three lasers undergo a very similar spectral evolution. Above threshold, the laser remains single mode for a substantial range of current until a clear instability threshold is reached, at which a 1-mA increase in current results in the appearance of new lasing modes. The new frequencies appear as symmetric sidebands on the primary lasing frequency, with a separation that is many integer multiples of the FSR. The sideband spacing $\delta\omega_{\text{sb}}$ and pumping $J_{\text{sb}}$ at the sideband instability threshold are given in
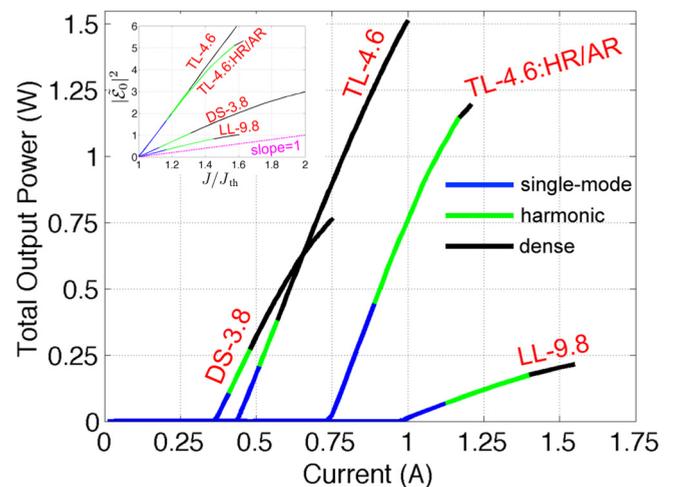


FIG. 2. Total power output of each QCL (from both facets) vs current, color coded to indicate the range over which the laser operates in a single mode (the region following threshold), harmonic state (the middle region), or dense state (the highest-power region). Inset: the intracavity power normalized to the saturation intensity (calculated from the measured output power and the best estimates for $\kappa$, $T_1$, $T_2$, and the facet reflectivities) is plotted vs $J/J_{\text{th}}$.
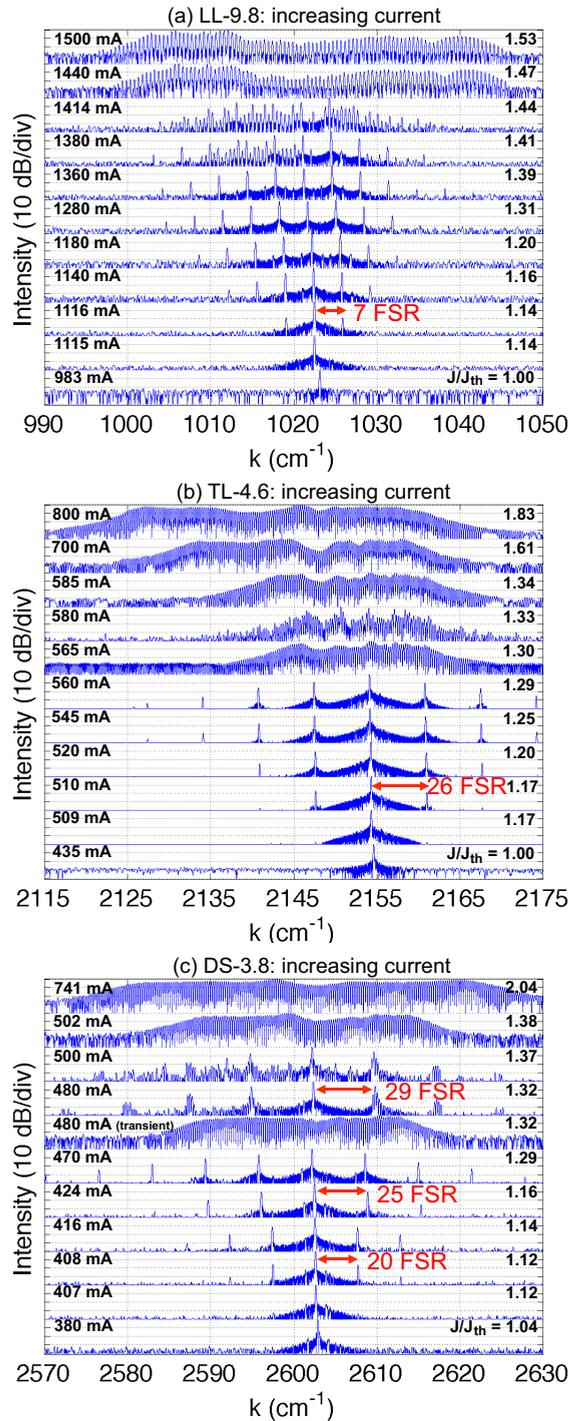
FIG. 3. Spectra of the three uncoated QCLs (a) LL-9.8, (b) TL-4.6, and (c) DS-3.8 as the current is incremented, starting from below threshold.

Table I for each device. Taking LL-9.8 as a first example, at $J_{sb}/J_{th} = 1.14$ a pair of equal-amplitude sidebands separated by 7 FSR from the primary mode suddenly rise out of the noise floor to an intensity 20 dB weaker than the primary mode. As the current increases further, higher-order sidebands appear that preserve the initial spacing, eventually yielding a spectrum at $J/J_{th} = 1.39$ of 11 modes, each separated by 7 FSR from its nearest neighbors. We refer to a spectrum of

modes separated by multiple FSR as a harmonic state. Above $J/J_{th} = 1.39$, interleaving modes incommensurate with the harmonic spacing begin to appear. At $J/J_{th} = 1.47$, there is another sudden transition at which all adjacent FP modes are populated; we refer to this as a "dense" state, and it persists for all higher currents. For device TL-4.6, sidebands with a separation of 26 FSR from the primary mode appear at $J_{sb}/J_{th} = 1.17$, and the transition to the dense state occurs at $J/J_{th} = 1.30$. (In a second device nominally identical to TL-4.6, sidebands separated by 13 FSR appeared at $J_{sb}/J_{th} = 1.19$ and the dense state appeared at $J/J_{th} = 1.31$.) For device DS-3.8, the sideband separation is 20 FSR at $J_{sb}/J_{th} = 1.12$. As the current is increased, the sideband spacing displays a sudden jump from 20 to 25 FSR. At $J/J_{th} = 1.32$ the laser jumps to a dense state for somewhere between a few seconds and a minute before returning to a "noisy" harmonic state: one with prominent harmonic peaks but many incommensurate modes populated as well. At $J/J_{th} = 1.38$, the dense state appears again, and this time persists for all higher currents.

When the spectral evolution measurement is repeated many times for one device, starting from below threshold and incrementing the current, we find that the instability threshold $J_{sb}$ and sideband spacing $\delta\omega_{sb}$ are always the same. As the current is increased past $J_{sb}$, there can be slight variations from one experiment to another. For example, the jump from 20 to 25 FSR in DS-3.8 does not always occur at the exact same current, but predictably within a range of about 20 mA. The same is true of the transition to the dense state.

After the laser enters the dense state, we decrease the current slowly and observe a remarkable hysteresis in the spectral evolution, shown in Fig. 4. In LL-9.8, the dense state persists all the way until $J/J_{th} = 1.01$, when the single mode finally reappears. In TL-4.6, the dense state gives way to a single mode at 2148 cm$^{-1}$ at $J/J_{th} = 1.10$, and then at $J/J_{th} = 1.01$ jumps to a single mode at 2155 cm$^{-1}$, which is the same mode observed at threshold when the current is ramped up in Fig. 3(b). In DS-3.8, a noisy harmonic state appears at $J/J_{th} = 1.38$, then the laser returns to the dense state at $J/J_{th} = 1.14$, and the single-mode state reappears at $J/J_{th} = 1.05$. We emphasize the general observation for all three devices that the clean harmonic state cannot be recovered once the current has been increased far into the dense state regime. (If the ramp down is begun from a current not too much larger than the one at which the laser enters the dense state, harmonic states can reappear.) Additionally, there is no spectral hysteresis in the immediate vicinity of the instability threshold; for example, if DS-3.8 is toggled between 407 and 408 mA, then the spectrum simply toggles between the two spectra shown in Fig. 3(c), and the same is true for LL-9.8 and TL-4.6.

One might expect sudden changes in the emission spectrum to be accompanied by changes in the output power and voltage. Since we can more sensitively measure the voltage than the output power, we plot in Fig. 5 a portion of the IV curve of DS-3.8. When starting below threshold and increasing the current, the voltage of the laser decreases (negative differential resistance) when the noisy harmonic state transitions to the dense state at 523 mA. [Note that in the spectra shown in Fig. 3(c), this transition occurred at 502 mA.] The IV curve also exhibits a hysteresis correlated with the spectral hysteresis: as the current is decreased after reaching the dense state, the laser
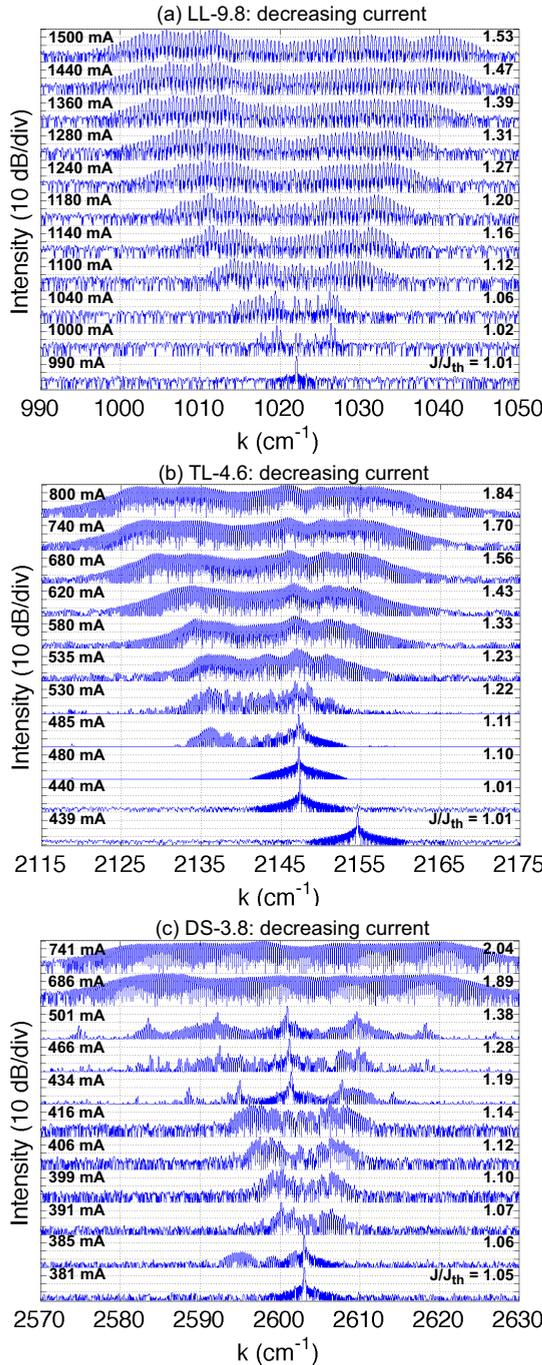
FIG. 4. Spectra of the three uncoated QCLs (a) LL-9.8, (b) TL-4.6, and (c) DS-3.8 as the current is decremented, starting from the current reached at the end of the upward current ramp shown in Fig. 3.
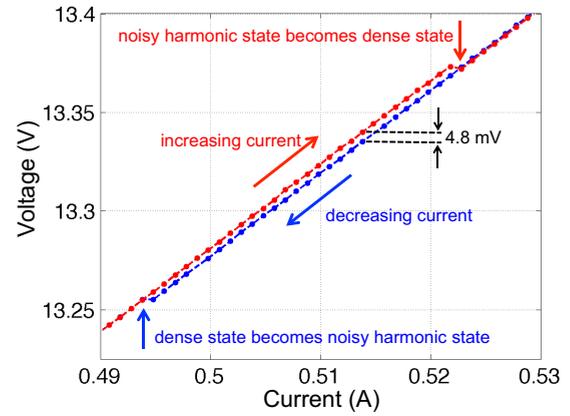


FIG. 5. The IV curve of DS-3.8 exhibits a hysteresis as the current is increased (red, upper curve) and decreased (blue, lower curve). The hysteresis is correlated with the transition from the noisy harmonic state to the dense state on the ramp up, and from the dense state to the noisy harmonic state on the ramp down.

46 FSR, much larger than any spacing seen previously, and (2) the harmonic regime persists over a much larger range of output power, from $J/J_{th} = 1.22$ to $1.60$, than it does in the uncoated devices, as seen by the color coding in Fig. 2. (A second device, nominally identical to TL-4.6:HR/AR, developed sidebands with a spacing of 48 FSR at $J_{sb}/J_{th} = 1.18$ and also remained in the harmonic state over a large current range.) As the current is decreased, shown in Fig. 6(b), the clean harmonic state with one pair of sidebands reappears at 885 mA, which is quite close to the instability threshold of 880 mA found when the current is ramped upwards. The single-mode state reappears at $J/J_{th} = 1.20$.

To better understand the coherence properties of a multimode state, it is typical to look at the width of the intermode radio-frequency beat note generated on a photodetector. We present some measurements of the beat note in the dense state in Appendix C. We find a range of interesting phenomena at various currents [39], including narrow beat notes, multiple closely spaced beat notes, and broad beat notes. However, the focus of this paper is not on the coherence properties of the dense state [50], but rather to understand the transition from the single mode to the harmonic state. Unfortunately, we cannot perform beat note measurements of the harmonic state because the smallest observed beat frequency is greater than 100 GHz, larger than the electrical bandwidth of any mid-infrared photodetector. The observed spectra of the harmonic state strongly suggest, however, that the modes are phase locked with one another through a nonlinear interaction because it would be difficult to explain the mode skipping if each lasing frequency acted as an independent oscillator. In the following theory section, we assume that the primary mode is coherent with the two sidebands and find that the consequences are consistent with our observations. The theory predicts that the uncoated devices emit FM waveforms, and suggests that the HR/AR device emits an AM waveform. In future work, second-order autocorrelation measurements are needed to experimentally verify this prediction.

remains in the dense state. For the same current, the voltage is 4.8 mV lower in the dense state than in the noisy harmonic state. (Accordingly, the output power is slightly higher in the dense state.) At 494 mA, the dense state transitions to the noisy harmonic state, and the two voltage curves overlap again.

Lastly, we present in Fig. 6(a) the spectral evolution of TL-4.6:HR/AR as the current is incremented. The behavior of this device is different from the uncoated devices in two significant ways: (1) the sidebands appear with a separation of
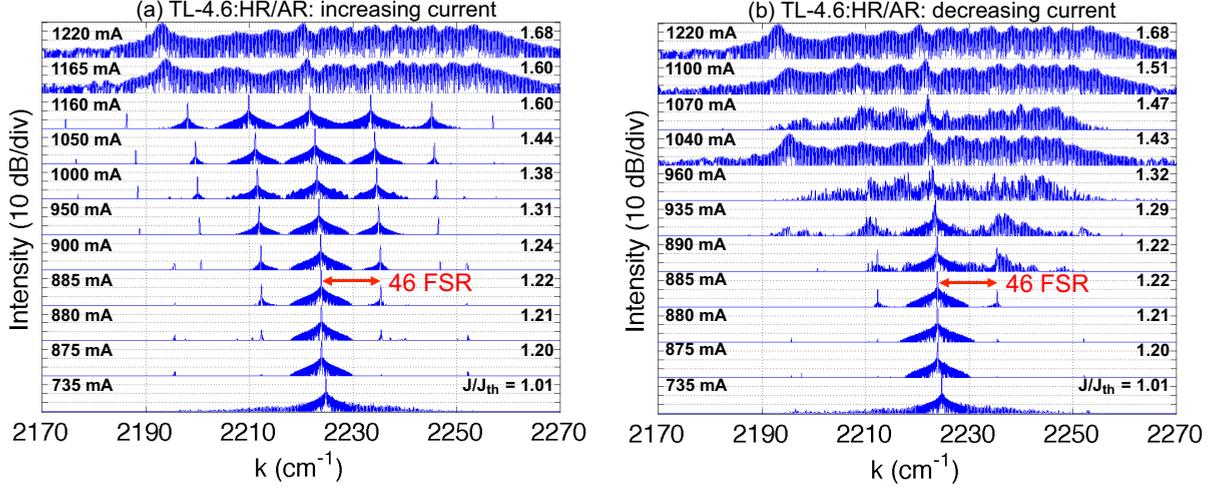
FIG. 6. Spectra of TL-4.6:HR/AR as the current is (a) incremented, starting from below threshold, and (b) decremented, starting from the current reached at the end of the upward current ramp in (a).

## III. THEORY

The instability threshold is characterized by the appearance of symmetric sidebands on the primary lasing mode. Our goal is to theoretically explain the frequency separation of the sidebands and the pump power at which they first appear. We begin with the general framework: the Maxwell-Bloch equations for a two-level system and the spatial mode expansion of a laser cavity. Then, we address the single-mode solution of the laser to determine how the primary mode and the population grating (PG) evolve with increasing pumping. Next, we understand how a two-level system responds to a weak field at a frequency different from that of the primary mode: the population pulsation (PP). Finally, we will combine these two ingredients, the PG and the PP, to explain the instability threshold. We find that the PG provides an unclamped Lorentzian contribution to the gain of the sidebands, which is responsible for the low instability threshold. The PP reshapes the gain, suppressing nearby sidebands and enhancing more distant ones, and is responsible for the observed multiple-FSR sideband separation. Interestingly, we find that depending on the relative contributions of the PG and the PP to the gain, the laser can emit either an FM or AM waveform at the instability threshold.

### A. General framework

We model the lasing transition as a two-level system, or a quantum dipole, subject to the electric field

$$E(t) = \mathcal{E}(t)e^{i\omega_0 t} + \text{c.c.}, \tag{1}$$

The response is characterized by the population inversion $w$ (positive when inverted) and the off-diagonal element of the density matrix $\sigma$, which in turn obey the Bloch equations (in the rotating-wave approximation) [51]

$$\dot{\sigma} = \left(i\Delta - \frac{1}{T_2}\right)\sigma + \frac{i\kappa}{2}w\mathcal{E}, \tag{2}$$

$$\dot{w} = i\kappa(\mathcal{E}^*\sigma - \mathcal{E}\sigma^*) - \frac{w - w_{\text{eq}}}{T_1}, \tag{3}$$

where $\Delta = \omega_{\text{ba}} - \omega_0$ is the detuning between the field and the resonant frequency $\omega_{\text{ba}}$ of the two-level system, $T_1$ is the gain recovery time, $T_2$ is the dephasing time, $\kappa \equiv 2d/\hbar$ is the coupling constant where $d$ is the dipole matrix element (assumed to be real) and $\hbar$ is Planck's constant, and $w_{\text{eq}}$ is the "equilibrium" population inversion that the system would reach in the absence of photons, determined by the pumping. (Note that we have defined $T_1$ to be the gain recovery time, which in QCLs is distinct from the upper-state lifetime $T_{\text{up}}$ due to the nature of electron transport in the active region. See Appendix D for a discussion of this point.) We write the macroscopic polarization $P$ (dipole moment per volume) as

$$P(t) = \mathcal{P}e^{i\omega_0 t} + \text{c.c.}, \tag{4}$$

where $\mathcal{P} = Nd\sigma$, and $N$ is the volume density of dipoles.

A characteristic of the two-level medium that will appear often is the "Beer rate"

$$\bar{\alpha} = \frac{Nd^2 T_2 \omega_{\text{ba}} c\sqrt{\mu/\epsilon}}{\hbar}, \tag{5}$$

with dimensions of frequency, which is related to the more familiar Beer coefficient $\alpha$ (with dimensions of inverse length) that appears in Beer's law of absorption by $\bar{\alpha} = \alpha c$. The Beer rate gives the amount of loss when the material is in its ground state ($w = -1$), and also the maximum amount of gain when the material is fully inverted ($w = 1$). We adopt the convention of [52] and assume our dipoles to be embedded in a host medium of permittivity $\epsilon$ and permeability $\mu$. The speed of light $c = 1/\sqrt{\epsilon\mu}$ also denotes the value in the background host medium.

In the laser cavity, the field envelopes vary in space and time. One approach is to numerically solve the full spatiotemporal Maxwell-Bloch equations [53,54]. To obtain analytical results, we follow a common approximation and decouple the spatial and temporal dependencies [55], writing the field as

$$E(z,t) = \sum_{m=-,0,+} \Upsilon_m(z)\mathcal{E}_m(t)e^{i\omega_m t} + \text{c.c.}, \tag{6}$$

where $\omega_0$ is the primary mode frequency and the sideband frequencies are $\omega_\pm = \omega_0 \pm \delta\omega$. These three frequencies are

cold-cavity resonant frequencies, and are equidistant from one another because we have assumed zero GVD. We will henceforth assume that the primary mode $\omega_0$ lases at the resonant frequency of the two-level system, so $\Delta = 0$. This is a reasonable approximation if the FSR is much smaller than the gain bandwidth. These two assumptions, GVD = 0 and $\Delta = 0$, simplify later mathematical formulas considerably and allow for easier understanding of the essential physics. The full theory without these assumptions is included in Appendix F. The spatial modes $\Upsilon_m(z)$ are determined by the cavity geometry, and do not vary in time. We assume a linear FP cavity with two end mirrors of unity reflectivity as shown in Fig. 1(d), so that the spatial modes are given by

$$\Upsilon_m(z) = \sqrt{2}\cos(k_m z), \qquad (7)$$

where $k_m$ is an integer multiple of $\pi/L$ and $L$ is the length of the cavity. (The simpler case of the ring cavity is included in Appendix G.) The mirror loss $\ln(1/\sqrt{R_1 R_2})/L$ is included in the total optical losses of the cavity $\bar{\ell}$. The assumption of perfect reflectivity simplifies the problem in two important ways: the spatial functions $\Upsilon_m(z)$ are orthogonal, and they do not change shape as the pumping increases. This assumption turns out to be quite good even for semiconductor lasers with facet reflectivities around 0.25. The approximation breaks down for our HR/AR coated QCL, so the theory will only be directly applied to the uncoated lasers, but the implications of the theory for the HR/AR device will be qualitatively discussed.

### B. Population grating

The threshold inversion is given by the ratio of the optical loss rate to the Beer rate, $w_{th} = \bar{\ell}/\bar{\alpha}$. We define the pumping parameter $p \equiv w_{eq}/w_{th}$. When $p = 1$, the primary mode begins to lase at the frequency $\omega_0 = \omega_{ba}$. As the pumping $p$ is increased, the field and inversion can be solved for by the method of [33] as shown in Appendix E, the main results of which are stated here. We account for the population grating, but not the coherence grating which has been incorporated in recent work [56]. The primary mode $\tilde{\mathcal{E}}_0$ grows according to

$$|\tilde{\mathcal{E}}_0|^2 = \frac{p-1}{1+\gamma_D/2}, \qquad (8)$$

where we have defined the dimensionless primary mode amplitude $\tilde{\mathcal{E}}_0$ by normalizing by the saturation amplitude $\tilde{\mathcal{E}}_0 \equiv \kappa\sqrt{T_1 T_2}\mathcal{E}_0$. The diffusion parameter $\gamma_D$ is given by $\gamma_D = (1 + 4k_0^2 D T_{up})^{-1}$, where $D$ is the lateral diffusivity of the excited-state electrons and $T_{up}$ is the upper-state lifetime. The parameter $\gamma_D$ ranges from 0 (for infinite mobility) to 1 (for zero mobility). The population inversion in the presence of the primary mode $w_0(z)$ varies with $p$ as

$$w_0(z) = w_{th}\left[1 + \frac{\gamma_D}{2}\frac{p-1}{1+\gamma_D/2} - \gamma_D\frac{p-1}{1+\gamma_D/2}\cos(2k_0 z)\right]. \qquad (9)$$

Equations (8) and (9) are valid to first order in the primary mode intensity $|\tilde{\mathcal{E}}_0|^2$ or, equivalently, $p - 1 \ll 1$. Note that for zero diffusion ($\gamma_D = 1$), the slope efficiency of the laser is two thirds that of the infinite diffusion ($\gamma_D = 0$) case [57]. This

is because for infinite diffusion, the inversion is uniformly pinned to $w_{th}$ above threshold. For finite diffusion, as the pumping increases the population grating grows in amplitude. At the same time, the average value of the inversion increases, indicating that the inversion is not being converted into photons as efficiently as it could be if the electrons could diffuse from the field nodes to the antinodes. In this derivation, we have assumed that the pump parameter $p$ is constant along the length of the laser; in an efficient electrically pumped laser more current will flow to the field antinodes, which will reduce the amplitude of the population grating.

In principle, one can extract $\gamma_D$ from measurements of the slope of $|\tilde{\mathcal{E}}_0|^2$ versus $p$, which should be between zero and one. The inset of Fig. 2 shows $|\tilde{\mathcal{E}}_0|^2$ versus $J/J_{th}$. All curves have a slope greater than one, which suggests that $J/J_{th}$ is an underestimate of $p$. (See Appendix A for how $|\tilde{\mathcal{E}}_0|^2$ is determined from the measurements, and how a nonzero transparency current causes $J/J_{th}$ to underestimate $p$.) Therefore, more characterization is needed to extract $\gamma_D$ from the measurements.

### C. Population pulsation

To understand the population pulsation, we can ignore the spatial dependence of the intracavity field and consider only a single two-level system subject to an applied field

$$E(t) = \sum_{m=-,0,+} \mathcal{E}_m(t)e^{i\omega_m t} + \text{c.c.} \qquad (10)$$

Since we will later be interested in calculating the stability of the sidebands, the amplitudes $\mathcal{E}_\pm$ should be thought of as infinitesimal perturbations; as such, our entire treatment retains only terms to first order in the sideband amplitudes $\mathcal{E}_\pm$. Full details of the calculation are in Appendix F. We write the total polarization as

$$P(t) = \sum_{m=-,0,+} \mathcal{P}_m(t)e^{i\omega_m t} + \text{c.c.} \qquad (11)$$

The polarization at the sidebands can be calculated using Eqs. (2) and (3) [58], which gives

$$\mathcal{P}_+ = \frac{i\epsilon}{\omega_{ba}}\bar{\alpha}w_0\left[\frac{\mathcal{E}_+}{[1+i\delta\omega T_2]} + \Lambda\tilde{\mathcal{E}}_0\tilde{\mathcal{E}}_0^*\mathcal{E}_+ + \Lambda\tilde{\mathcal{E}}_0\tilde{\mathcal{E}}_0\mathcal{E}_-^*\right], \qquad (12)$$

$$\mathcal{P}_- = \frac{i\epsilon}{\omega_{ba}}\bar{\alpha}w_0\left[\frac{\mathcal{E}_-}{[1-i\delta\omega T_2]} + \Lambda^*\tilde{\mathcal{E}}_0\tilde{\mathcal{E}}_0^*\mathcal{E}_- + \Lambda^*\tilde{\mathcal{E}}_0\tilde{\mathcal{E}}_0\mathcal{E}_+^*\right], \qquad (13)$$

where

$$\Lambda = \frac{-(1+i\delta\omega T_2/2)}{[(1+i\delta\omega T_1)(1+i\delta\omega T_2)^2 + (1+i\delta\omega T_2)|\tilde{\mathcal{E}}_0|^2]} \qquad (14)$$

and $w_0$ is the saturated population inversion $w_0 = w_{eq}/(1 + |\tilde{\mathcal{E}}_0|^2)$.

The polarization at each sideband is neatly divided into three contributions. Taking $\mathcal{P}_+$ as an example, the first term in Eq. (12) is the Lorentzian contribution that the sideband generates due to the linear susceptibility of the dipole. The second and third terms are nonlinear contributions due to the

PP at frequency $\delta\omega$: a self-mixing term of the sideband with the primary mode, and a cross-mixing term of the other sideband with the primary mode. The frequency-dependent portion of the nonlinear susceptibility is $\Lambda$, which is a dimensionless function of the sideband detuning, the time constants of the two-level system, and the primary mode intensity $|\tilde{\mathcal{E}}_0|^2$. From the field and the induced polarization, we can calculate the total power density generated $\langle -E\dot{P}\rangle$. The quantity that most interests us is the gain $\bar{g}$ (with dimension of frequency) seen by each sideband, defined as the power generated at the sideband's frequency, divided by the energy density of the exciting sideband field.

To develop a feel for the parametrically generated polarization and the resulting gain, we consider two instructive cases. In both cases, we take the sidebands to have equal magnitudes $|\mathcal{E}_+| = |\mathcal{E}_-|$, but choose the phases of the sidebands to give rise to an AM waveform in one case and a constant-intensity FM waveform in the other case, as shown in Fig. 1(b). The gain of each sideband is found to be

$$\bar{g} = \bar{\alpha}w_0\left[\frac{1}{1+(\delta\omega T_2)^2} + \mathrm{Re}(\Lambda)|\tilde{\mathcal{E}}_0|^2 \times \begin{cases} 2 & \text{(AM)} \\ 0 & \text{(FM)} \end{cases}\right].$$
(15)

The first term is the Lorentzian contribution to the gain, and the second term is the parametric gain due to the PP. The factors of 2 and 0 come from the constructive or destructive addition, respectively, of the self-mixing and cross-mixing terms to the nonlinear polarization. Equivalently, one can say that the constant-intensity FM field does not create a PP, and accordingly experiences no parametric gain. The parametric gain of the AM field is proportional to $\mathrm{Re}(\Lambda)$ and to the primary mode intensity $|\tilde{\mathcal{E}}_0|^2$. [We note that one can quickly derive the original RNGH instability for a traveling-wave laser from Eq. (15), which is shown in Appendix G.] By expanding $\Lambda$ in Eq. (14) in powers of $|\tilde{\mathcal{E}}_0|^2$, it becomes clear that the PP interaction can be expressed in the perturbative expansion of traditional nonlinear optics as a third-, fifth-, seventh-order, etc., nonlinearity. We will later calculate the instability threshold in the limit of small primary mode intensity, and are therefore interested in the lowest-order nonlinearity. We obtain $\chi^{(3)}$, the dimensionless frequency-dependent portion of the third-order PP nonlinear susceptibility, by evaluating $\Lambda$ at $|\tilde{\mathcal{E}}_0|^2 = 0$:

$$\chi^{(3)} = \frac{-(1+i\delta\omega T_2/2)}{(1+i\delta\omega T_1)(1+i\delta\omega T_2)^2}.$$
(16)

To better elucidate the nature of the PPs, the magnitude, phase, and real part of $\chi^{(3)}$ are plotted in Fig. 7 as a function of the sideband detuning, for three different values of $T_1/T_2$. At low frequencies $\delta\omega$, the population inversion has no difficulty following the modulation of the field, which has two consequences: the amplitude of the PPs is large, and the PP is $\pi$ out of phase with the intensity modulation of the exciting field. This can be understood simply in terms of rate equations: when the field is stronger, the stimulated emission rate is larger, and the population inversion is therefore smaller. This scenario, higher inversion when the intensity is lower and lower inversion when the intensity is higher, is less efficient at extracting power from the two-level system relative to the case of monochromatic or FM excitation; mathematically, this
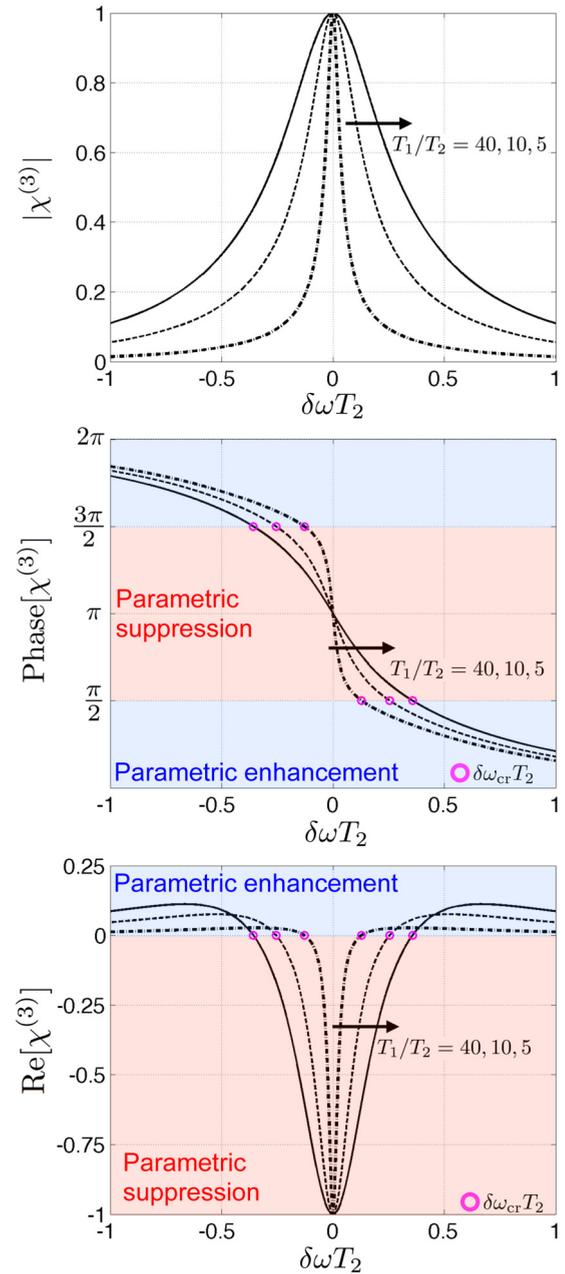


FIG. 7. The magnitude, phase, and real part of $\chi^{(3)}$ are plotted vs $\delta\omega T_2$ for three different ratios $T_1/T_2 = 40, 10, 5$. The parametric gain seen by the sidebands is determined by $\mathrm{Re}[\chi^{(3)}]$. Low-frequency PPs lead to a parametric suppression of the gain. For the gain to be parametrically enhanced, $\delta\omega$ must be large enough that the inversion can no longer follow the intensity in antiphase; in other words, the phase of $\chi^{(3)}$ must be between $-\pi/2$ and $\pi/2$.

is described by a parametric gain (determined by the real part of $\chi^{(3)}$) that is negative. We refer to this effect as parametric suppression: a low-frequency PP reduces the gain of each sideband. As $\delta\omega$ increases, the inversion can no longer as easily follow the intensity modulation, so the amplitude of the PPs decreases and the phase of $\chi^{(3)}$ decreases from $\pi$. For large enough $\delta\omega$, the phase of $\chi^{(3)}$ decreases below $\pi/2$, at which point $\mathrm{Re}[\chi^{(3)}]$ becomes positive. We refer to this effect as parametric enhancement: a high-frequency PP increases

the gain seen by each sideband. The crossing frequency $\delta\omega_{\mathrm{cr}}$ which separates the low-frequency suppression regime and high-frequency enhancement regime is given by

$$\delta\omega_{\mathrm{cr}} T_2 \approx \sqrt{\frac{2/3}{T_1/T_2}}, \qquad (17)$$

where we have made the approximation $T_1/T_2 \gg 1$, valid for QCLs. The regions of parametric suppression and enhancement are highlighted in the plots of the phase and real part of $\chi^{(3)}$ in Fig. 7. Finally, at very large $\delta\omega$ the parametric gain approaches zero (from above) because the beat note becomes too short for the inversion to follow and the amplitude of the PP approaches zero.

It is worth pointing out that in the weak-field limit $|\tilde{\mathcal{E}}_0|^2 \ll 1$ that we are interested in, $\delta\omega_{\mathrm{cr}}$ has no relation to the Rabi frequency $\Omega_R$ induced by the primary mode

$$\Omega_R T_2 = \frac{|\tilde{\mathcal{E}}_0|}{\sqrt{T_1/T_2}}. \qquad (18)$$

The Rabi frequency of course varies with the primary mode amplitude, while $\delta\omega_{\mathrm{cr}}$ is independent of $\tilde{\mathcal{E}}_0$ in the weak-field limit. By comparing the factors $\sqrt{2/3}$ and $|\tilde{\mathcal{E}}_0|$ in the numerators of Eqs. (17) and (18), it is clear that in the limit $|\tilde{\mathcal{E}}_0|^2 \ll 1$, $\delta\omega_{\mathrm{cr}}$ will always be greater than the Rabi frequency. Thus, the reason for the parametric enhancement when $\delta\omega > \delta\omega_{\mathrm{cr}}$ should simply be ascribed to the fact that at high PP frequency, the phase lag between the population inversion and the field intensity becomes appropriate for gain rather than absorption.

### D. Instability threshold

Now, we can ask the following question: What happens to the single-mode laser solution when it is perturbed by a weak sideband field? The source of the perturbation could be spontaneous emission, or even spontaneous parametric down conversion of two primary mode photons into two sideband photons [59]. Our goal is to calculate the gain of the sideband modes averaged over the length of the cavity. The instability threshold is reached when the sideband gain equals the loss. Although our instability analysis will not tell us about the steady state reached by the sidebands, one reasonable possibility is that the sidebands begin to lase, as seen in the experimental spectra.

To determine the sideband gain, we start with the polarization in Eqs. (12) and (13) and account for the position dependence by replacing $\mathcal{E}_m(t)$ with $\mathcal{E}_m(t)\Upsilon_m(z)$, and $w_0$ with $w_0(z)$ from Eq. (9). In keeping with our approximation to order $|\tilde{\mathcal{E}}_0|^2$, we replace $\Lambda$ with $\chi^{(3)}$. The position-dependent polarization is then inserted as the source term in Maxwell's wave equation. From here, the calculation follows the same steps as the instability analysis done for Kerr microresonators [55], and is detailed in Appendix G. After making the slowly varying envelope approximation and projecting the equation onto each of the orthonormal spatial modes, one finds a first-order differential equation for each sideband amplitude. Unlike the earlier example where we hand picked the phases of the sidebands to study the effect of an AM and FM field, here the AM and FM sideband configurations emerge organically

as the two "natural modes" of the system of two sideband equations. The natural modes [14] are the configurations of the three-wave field for which the relative phases of the fields are preserved as time evolves; in other words, an AM field remains AM, and an FM field remains FM. (In the general case of nonzero $\Delta$ and GVD, the natural modes can be a superposition of AM and FM.) The gain of the AM and FM natural modes is given by

$$\frac{\bar{g}}{\bar{\ell}} = \frac{1 + \frac{\gamma_D}{2}\frac{p-1}{1+\gamma_D/2}}{1 + (\delta\omega T_2)^2}$$

$$+ \operatorname{Re}[\chi^{(3)}]\frac{p-1}{1+\gamma_D/2} \times \begin{cases} \Gamma_{\mathrm{self}} + \Gamma_{\mathrm{cross}} = \frac{3}{2} & \text{(AM)} \\ \Gamma_{\mathrm{self}} - \Gamma_{\mathrm{cross}} = \frac{1}{2} & \text{(FM)} \end{cases}, \qquad (19)$$

where the $\Gamma$'s are longitudinal spatial overlap factors

$$\Gamma_{\mathrm{self}} = \frac{1}{L}\int_0^L dz\, |\Upsilon_0(z)|^2 |\Upsilon_\pm(z)|^2 = 1, \qquad (20)$$

$$\Gamma_{\mathrm{cross}} = \frac{1}{L}\int_0^L dz\, \Upsilon_0(z)^2 \Upsilon_-^*(z) \Upsilon_+^*(z) = 1/2. \qquad (21)$$

By comparing the standing-wave sideband gain in Eq. (19) to the sideband gain of a single two-level system in Eq. (15), we see that the cavity introduces two modifications. First, the Lorentzian gain contribution increases with $p$; this unclamped gain is a direct result of the PG that develops in the presence of nonzero $\gamma_D$. Secondly, the partial overlap of the sideband spatial modes $\Upsilon_+$ and $\Upsilon_-$ results in partial (rather than complete) interference of the self-mixing and cross-mixing contributions to the gain. To understand this, note from Fig. 1(a) that although the *emitted* waveform has equal-amplitude sidebands, *within* the cavity the plus and minus sidebands have unequal amplitudes at most positions $z$, as shown by the red and blue modes in Fig. 1(d). Therefore, the self- and cross-mixing contributions to the sideband polarization at each position $z$ cannot completely interfere, and the factors of $\frac{3}{2}$ (AM) and $\frac{1}{2}$ (FM) emerge after averaging over the full cavity length, as opposed to the factors of 2 and 0 in Eq. (15). Thus, even when the laser emits an FM waveform, there is still a parametric contribution to the gain due to the incomplete destructive interference of the PP within the cavity.

The instability occurs when $p$ reaches a value such that the sideband gain $\bar{g}$ in Eq. (19) equals the loss $\bar{\ell}$ for one particular sideband detuning $\delta\omega$. (As discussed previously, we assume the FSR is small so that a FP mode always exists very close to the unstable value of $\delta\omega$.) As $p$ increases, the incoherent Lorentzian gain increases, but the parametric gain either increases or becomes more negative depending on the sign of $\operatorname{Re}[\chi^{(3)}]$ (which depends on $\delta\omega$). The three parameters $T_1$, $T_2$, and $\gamma_D$ affect the relative importance of the incoherent and coherent gain terms, and depending on the values of these parameters, one of three different classes of instability can occur: the incoherent instability, FM instability, and AM instabilty. In Fig. 8, each type of instability is illustrated by plotting the sideband gain at the instability threshold, which we now explain.
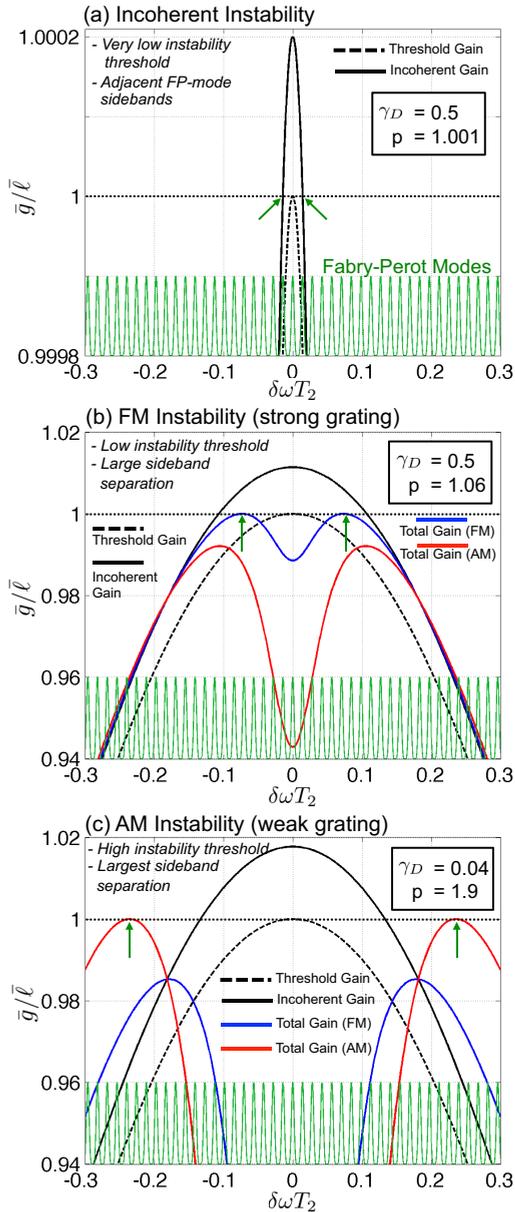
FIG. 8. Overview of the three different types of instabilities. (a) The incoherent instability relies only on the unclamped gain due to the PG, and occurs when parametric effects can be neglected. (b) The FM instability occurs when the gain due to a strong PG, despite the parametric suppression of the gain of small-detuned sidebands, allows the less-suppressed FM sidebands to reach threshold. (c) The AM instability occurs when the gain due to a weak PG, together with the parametric enhancement of the gain of large-detuned sidebands, allows the more strongly enhanced AM sidebands to reach threshold. In (b) and (c), the value $T_1/T_2 = 20$ was used. The FP modes (green, bottom of each panel) are not associated with the ordinate, and simply provide a sense of the mode spacing.

### 1. Incoherent instability

The parametric gain can often be ignored. If $T_1$ is large enough, the interesting features of $\text{Re}[\chi^{(3)}]$ all occur for sideband detunings less than 1 FSR, and so the parametric gain will be nearly zero for all values of $\delta\omega$ greater than 1 FSR. This

is the case for diode lasers, where PPs are significant up to a few GHz ($T_1 \approx 1$ ns), while the FSR is around 100 GHz. Thus, only the incoherent gain term in Eq. (19) matters (although it is not a Lorentzian for band-gap lasers). As $p$ increases beyond 1, the sideband gain increases but remains Lorentzian, so the sidebands that reach the instability threshold first will always be the FP modes immediately adjacent to the primary lasing mode [60]. In diode lasers, $\gamma_D$ is small ($\sim 10^{-4}$), so $p$ needs to be large before the second mode can appear.

The value of $\gamma_D = 0.5$ in Fig. 8(a) is typical of short-wave QCLs. We see that if coherent effects were negligible in QCLs, we would expect the sidebands to appear at $p = 1.001$, barely above threshold. The much higher instability threshold measured in the experiments, together with the observation that the sidebands do not appear at the nearest-neighbor FP modes of the primary mode, indicates that coherent effects play an essential role in the QCL instability.

### 2. FM instability

When the Lorentzian gain increases quickly with $p$ due to a strong PG, sidebands that fall within the parametric suppression regime, $\delta\omega < \delta\omega_{\text{cr}}$, can reach the instability threshold. This is counterintuitive: why should a sideband lase when the parametric interaction provides negative gain? The answer is that the Lorentzian gain favors sidebands with as small a separation as possible, and if the Lorentzian gain is large enough it can pull sidebands above threshold in spite of the negative contribution from the parametric gain. In this scenario, FM sidebands have a lower instability threshold than AM sidebands because the parametric contribution to the gain is *less negative* since $\frac{1}{2} < \frac{3}{2}$ in Eq. (19). Such a case is illustrated in Fig. 8(b) for $\gamma_D = 0.5$ and $T_1/T_2 = 20$. At $p = 1.06$, FM sidebands reach the instability threshold, while AM sidebands are too strongly suppressed to reach the instability. A key feature of the instability is that the unstable sideband will be several FSR away from the primary mode (provided that the FSR is small), while still satisfying $\delta\omega < \delta\omega_{\text{cr}}$.

### 3. AM instability

When the Lorentzian gain increases little with $p$ due to a weak PG, only sidebands that fall within the parametric enhancement regime $\delta\omega > \delta\omega_{\text{cr}}$ will be able to reach the instability. In this case, AM will have a lower instability threshold than FM because AM receives a larger parametric enhancement [since $\frac{3}{2} > \frac{1}{2}$ in Eq. (19)]. Such a case is illustrated in Fig. 8(c) for $\gamma_D = 0.04$ and $T_1/T_2 = 20$. At $p = 1.9$, AM sidebands reach the instability threshold while the FM sidebands are not sufficiently enhanced to reach the instability. Strictly speaking, $p = 1.9$ falls outside the region of validity of our perturbative treatment ($p - 1 \ll 1$), so the specific values in this plot are not exactly accurate, but the qualitative features are correct. The unstable sidebands satisfy $\delta\omega > \delta\omega_{\text{cr}}$, and so their separation will be even greater than for the FM instability. The original RNGH instability is precisely this AM instability, in a traveling-wave laser. For traveling waves, the Lorentzian gain is clamped at threshold regardless of the diffusion parameter, so the instability can only be reached by the parametric enhancement of AM sidebands.

To access both the FM and AM instability regimes experimentally, we need to tune the strength of the PG. The electron diffusivity can be reduced by lowering the temperature, and indeed temperature has a strong effect on the emission spectra of QCLs [33], although the effect is not yet well understood. In this work, we choose to manipulate the PG by adjusting the facet reflectivities. Increasing the disparity of the reflectivities of the two mirrors reduces the contrast of the standing wave because the wave traveling from the higher- to the lower-reflectivity facet becomes larger than the counterpropagating wave [61]. For a sufficiently large disparity, the incoherent gain contribution is small enough that the laser can only undergo the AM instability. In a practical sense, engineering the facet coatings allows one to transform a standing-wave cavity into more of a traveling-wave cavity. It is for this reason that we chose to study an HR/AR coated laser, where the AR coating has as low a reflectivity as current technology allows, to maximize the cavity asymmetry. The full mathematical treatment of mirrors with nonunity reflectivity is complicated by the fact that the spatial modes $\Upsilon_m(z)$ are no longer orthogonal, and also that that the $\Upsilon_m(z)$ and the longitudinal overlap factors $\Gamma$ vary with the pumping. This theory will be presented in future work.

## IV. DISCUSSION

In order for a mode to oscillate, it must satisfy two conditions: (1) the round-trip gain must equal the loss, and (2) the round-trip phase must equal a multiple of $2\pi$. Our theory in Sec. III has treated only the gain condition. The same approach was taken in the description of the original RNGH instability [10–14]; the underlying assumption is that the cavity modes are densely spaced, so that a pair of sidebands that satisfies the instability condition for the gain will always be "close enough" to two cavity modes that satisfy the phase condition. However, the experimental and theoretical developments of the last decade concerning optical parametric oscillation in externally pumped microresonators have shown that the phase condition has a large effect on the oscillation threshold and sideband spacing [55]. In microresonator experiments, the detuning between the external pump frequency and the center frequency of the cold-cavity mode is a degree of of freedom that must be precisely controlled to achieve the lowest possible instability threshold. In a laser, this detuning can not be experimentally controlled, but it most likely varies with the pumping in a deterministic manner and should be properly accounted for in a more complete theory. Furthermore, a parameter that has no analogy in microresonators is $\Delta$, the detuning between the lasing mode $\omega_0$ and the two-level resonance $\omega_{ba}$, which also varies with the pumping and is difficult to control in experiments. To precisely predict the instability threshold would require knowledge of both of these detunings, as well as the GVD.

At this stage, the simplest and most important application of the theory is to help determine whether the observed sidebands are parametrically enhanced or suppressed. Because the theory assumes end mirrors with unity reflectivity, we can only expect Eq. (19) to apply reasonably well to the uncoated QCLs. For each device, $\gamma_D$ is calculated using the theoretical value of $T_{up}$ (calculated from the band structure) and the diffusion constant

$D = 77$ cm$^2$/s [49], giving $\gamma_D = 0.4$ (DS-3.8), 0.49 (TL-4.6), and 0.93 (LL-9.8). For these large values of $\gamma_D$ the PG is strong, and we find from numerically solving Eq. (19) that the FM instability has a lower threshold than the AM instability, regardless of the value of $T_1$. In Appendix H, we show that the theory predicts sideband spacings $\delta\omega_{sb}$ that are consistent with the experimental observations, but underestimates the instability threshold $p_{sb}$. We attribute this discrepancy to the aforementioned detunings and GVD, as well as current inhomogeneity, that our theory neglects.

A more direct method to discriminate between the parametric suppression and enhancement regimes is to compare the observed sideband spacing $\delta\omega_{sb}$ to the crossing frequency $\delta\omega_{cr}$. If $\delta\omega_{sb} < \delta\omega_{cr}$, the sidebands are parametrically suppressed and therefore the FM instability has the lower threshold. Thus, we reason that the sidebands must be FM because the AM state would be an unstable equilibrium. Similarly, if $\delta\omega_{sb} > \delta\omega_{cr}$, the sidebands are parametrically enhanced so for them to be stable they must be AM. Notably, this reasoning depends only on the behavior of $\chi^{(3)}$ as a function of $\delta\omega$; it can therefore be applied to the HR/AR as well as the uncoated lasers because we do not need to understand the specifics of the PG. To calculate $\delta\omega_{cr}$ from Eq. (17), we use our measured values of $T_2$ but still need an estimate for the gain recovery time $T_1$. Pump-probe experiments [62,63] and theory [64] have shown that $T_1$ is around 2 ps. From Eq. (17), we see that $\delta\omega_{cr}$ decreases with increasing $T_1$, so if we take $T_1 = 3$ ps as a generous upper bound on the gain recovery time, we establish a lower bound of $\delta\omega_{cr}$ at 2270 GHz (DS-3.8), 1730 GHz (TL-4.6), and 1660 GHz (LL-9.8). The measured values of $\delta\omega_{sb}$ for each uncoated laser [977 GHz (DS-3.8), 1259 GHz (TL-4.6), and 642 GHz (LL-9.8)] are all substantially smaller than the lower bound on $\delta\omega_{cr}$. This is consistent with the prediction that the uncoated lasers have a lower FM instability threshold than AM threshold, and with these two results we are reasonably confident that the uncoated lasers emit parametrically suppressed FM sidebands. In stark contrast, TL-4.6:HR/AR exhibits a large sideband separation of $\delta\omega_{sb} = 2216$ GHz. If we use the accepted value of $T_1$ equal to 2 ps, we find $\delta\omega_{cr} = 2120$ GHz. The observed sideband spacing is slightly larger than $\delta\omega_{cr}$, suggesting the enhancement regime. While a smaller gain recovery time or nonperturbative calculation would raise $\delta\omega_{cr}$ slightly, this is our first hint that TL-4.6:HR/AR emits parametrically enhanced AM sidebands.

The difference in the range of intracavity power over which the harmonic state persists in the uncoated versus coated lasers, as shown in Fig. 2, is additional evidence that the uncoated devices operate in the suppression regime and the HR/AR device operates in the enhancement regime. Here, we propose a qualitative explanation of this feature. Consider a laser operating in the suppression regime. While the FM state is more stable than the AM state in this regime because it minimizes the amount of gain suppression, the FM state still pays a gain penalty by skipping over modes which would have a greater Lorentzian gain contribution. The dense state could therefore extract more gain from the inverted population, provided that the many modes are phased in such a way that minimizes the amplitude modulation throughout the cavity, thereby avoiding the gain suppression associated with such a

low-frequency one-FSR beat note. Indeed, the measurement of negative differential resistance shown in Fig. 5 shows that the dense state extracts more gain than the harmonic state. Furthermore, previous experiments have shown that the dense state emission is largely FM, not AM [37]. This could explain why the uncoated lasers only exhibit the harmonic state over a small range of current: the laser soon finds a way to transition from the parametrically suppressed FM state to the favored dense state. The fact that TL-4.6:HR/AR exhibits the harmonic state over a large current range suggests that the harmonic state is more stable than the dense state, which could only be true for a harmonic state in the parametric enhancement regime. At a sufficiently high current, when the spectral span of the harmonic state approaches the gain bandwidth, the dense state finally becomes favored for its ability to lase on adjacent modes and extract more incoherent gain, despite no longer benefitting from the parametric enhancement.

We have argued that the dense state is more favored than the FM harmonic state. If this is the case, why do the uncoated QCLs choose to emit a harmonic state at all, and not simply jump from the single-mode state to the dense state as the current is increased? In fact, the spectral hysteresis shown in Fig. 4 proves that the dense state *is* the favored lasing state down to barely above threshold. However, this state can only be reached by decreasing the current after the laser has already entered the dense state at high current. When the laser starts in a single-mode state and the current is increased, there is clearly a barrier that prevents the transition to the dense state. In general, introducing noise allows a system to overcome energy barriers and explore a larger volume of its state space. It is likely that delayed optical feedback serves as such a noise source, and explains why it is difficult to observe the harmonic state when optical feedback is not eliminated.

## V. CONCLUSION

We have experimentally identified the single-mode instability of QCLs, which is characterized by the appearance of sidebands at FP modes not adjacent to the primary lasing mode. We have seen the behavior in QCLs at three different wavelengths, each based on a different active region design, and with both positive and negative GVD. Therefore, the phenomenon is a general feature of the electron-light dynamics of QCLs. The instability is reached due to the combined contributions of an incoherent gain due to the spatial population grating, and a coherent parametric gain due to the temporal population pulsations. Our theory predicts both an FM instability in situations where the incoherent gain contribution is large, and an AM instability when the incoherent gain contribution is small. This theory extends the RNGH instability of traveling-wave lasers to standing-wave laser cavities. To explore the possibility of AM emission, we coated the QCL facets with an HR and an AR coating to reduce the incoherent gain contribution; indeed, this modification substantially increases the sideband spacing, and it is likely that the waveform is AM. Following the first appearance of sidebands at the instability threshold, our measurements show that increasing the pumping generates more sidebands which preserve the initial spacing. This suggests that a cw QCL can self-start into a phase-locked harmonic frequency

comb. We have also placed our observations and theory within historical context, explaining the relation to optically pumped microresonators and the single-mode instability in traveling-wave lasers.

The future direction of this work is clear. At first, we can take guidance from the well-established understanding of microresonators and exploit their similarity with QCLs to further our understanding. The calculation of the instability threshold will be extended to account for GVD, so that the cold-cavity modes are not necessarily equidistant. We must also better understand the nature of the single-mode solution; specifically, how does its detuning from the resonant frequency $\omega_{ba}$, and also its detuning from the cold-cavity mode that it occupies, affect the nature of the instability threshold? We must account quantitatively for the nonunity facet reflections and the precise shape of the mode profile within the cavity. Experimentally, second-order autocorrelation experiments are needed to establish the temporal nature of these short-period waveforms.

## APPENDIX A: COMMENTS ON POWER VS CURRENT CURVES

In the single-mode regime, the intracavity intensity of the single mode determines the strength of the parametric interaction with the sideband fluctuations. Therefore, we would like to calculate the intracavity intensity from the measured output power. In the distributed loss approximation, the output power is given by

$$P_{\text{out}} = \frac{\alpha_m \langle E^2 \rangle L w h}{\sqrt{\mu/\epsilon}}, \tag{A1}$$

where $\alpha_m = \ln[1/(R_1 R_2)]/(2L)$, the length, width, and height of the cavity are $L$, $w$, and $h$, and the time-averaged intensity of the single mode is $\langle E^2 \rangle = 2|\mathcal{E}_0|^2$. We are assuming a uniform field intensity in the transverse dimensions, and therefore not worrying about the transverse overlap factor. We can rearrange this equation for the intracavity intensity

$$|\tilde{\mathcal{E}}_0|^2 \equiv \kappa^2 T_1 T_2 |\mathcal{E}_0|^2 = \frac{2d^2 T_1 T_2 \sqrt{\mu_0/\epsilon_0}}{\hbar^2 n_{\text{eff}} \alpha_m L w h} P_{\text{out}}. \quad (A2)$$

With this equation, we can convert the measured total output power of each laser into the intracavity intensity, using our measured values of the refractive index $n_{\text{eff}}$ and the dephasing time $T_2$, our best estimates for $d$ and $T_1$, and in the case of the HR/AR laser we have used $R_1 = 1$, $R_2 = 0.01$. The result is plotted as a function of $J/J_{\text{th}}$ in the inset of Fig. 2.

The theoretical formula for the intracavity intensity is

$$|\tilde{\mathcal{E}}_0|^2 = \frac{p-1}{1 + \gamma_D/2}, \quad (A3)$$

where $p \equiv w_{\text{eq}}/w_{\text{th}}$ is the pump parameter. We emphasize that $p$ is not the same as $J/J_{\text{th}}$. The slope of $|\tilde{\mathcal{E}}_0|^2$ versus $p$ is always between $\frac{2}{3}$ and 1, depending on the diffusion parameter $\gamma_D$. The reference line in the inset of Fig. 2 is drawn with a slope of one to indicate that each of the $|\tilde{\mathcal{E}}_0|^2$ versus $J/J_{\text{th}}$ curves has a slope greater than one. Therefore, we conclude that $J/J_{\text{th}}$ must underestimate $p$. One factor that contributes to this underestimation is the transparency current $J_{\text{trans}}$: a fixed amount of current that must be delivered to the active region simply to raise the inversion from a negative number to zero. To understand this simply, suppose that the equilibrium inversion scales like $w_{\text{eq}} \propto J - J_{\text{trans}}$, and that $J_{\text{trans}}$ remains a constant number at threshold and above. Then, the pump parameter $p \equiv w_{\text{eq}}/w_{\text{th}}$ is expressed in terms of $J$ as

$$p = \frac{J - J_{\text{trans}}}{J_{\text{th}} - J_{\text{trans}}}. \quad (A4)$$

For example, suppose that for a laser with $J_{\text{th}} = 500$ mA the harmonic state kicks in at 550 mA, or $J/J_{\text{th}} = 1.1$. If the transparency current was $J_{\text{trans}} = 250$ mA (in other words, half of the threshold current, which is reasonable for room-temperature QCLs), then the pump parameter at the harmonic state onset would be $p = (550–250)/(500–250) = 1.2$. Thus, $J/J_{\text{th}}$ underestimates $p$.

A more rigorous study is required to determine $J_{\text{trans}}$ for each laser, which can be done by measuring many lasers of the same active region but different lengths. Once $J_{\text{trans}}$ is known, the slope of $|\tilde{\mathcal{E}}_0|^2$ versus $p$ should fall between $\frac{2}{3}$ and 1 and in principle a value for $\gamma_D$ can be extracted, allowing one to quantify the amount of diffusion present.

### APPENDIX B: GROUP VELOCITY DISPERSION

It has recently been shown that the group velocity dispersion (GVD) of the QCL is an important parameter in determining the spectral properties of the dense state [39], and it is reasonable to assume that GVD plays a role in determining some properties of the harmonic state as well. We present the measured GVD of the three uncoated devices in Fig. 9. The GVD is extracted from a measurement of the subthreshold amplified spontaneous emission by the method of [48]. A more
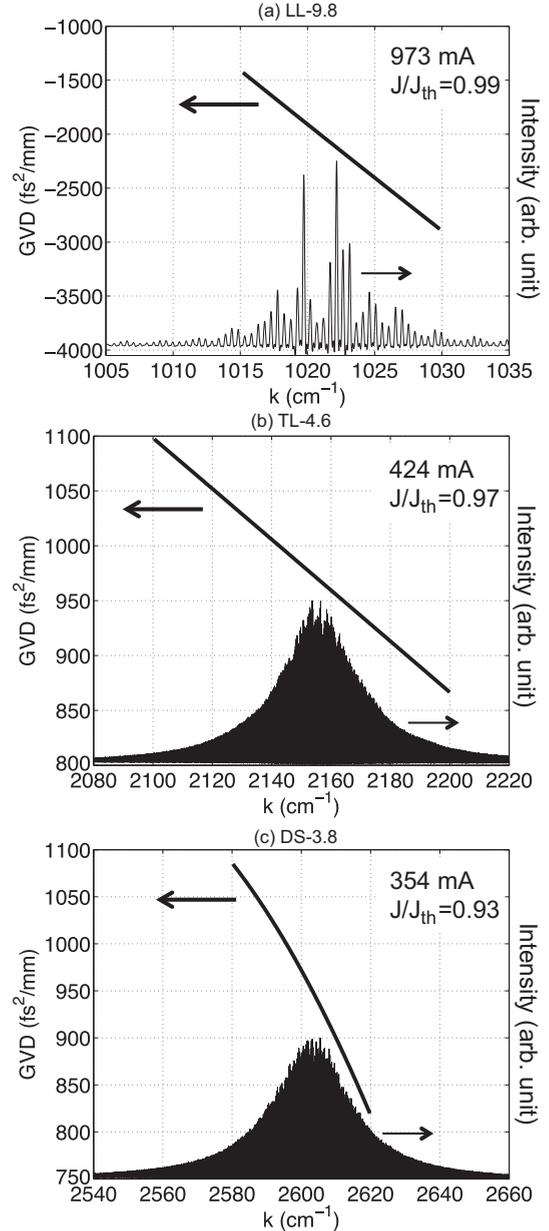


FIG. 9. The GVD is extracted from the subthreshold spectrum for (a) LL-9.8 at 973 mA, (b) TL-4.6 at 424 mA, and (c) DS-3.8 at 354 mA. The subthreshold spectra are plotted for reference on an arbitrary linear scale.

sensitive InSb detector (compared to HgCdTe) can be used for the lower wavelength devices TL-4.6 and DS-3.8, allowing the measurement to be performed further below threshold. This yields a broader spontaneous emission spectrum, and a larger bandwidth over which the GVD can be extracted. Devices TL-4.6 and DS-3.8 both have positive GVD, around 950 fs$^2$/mm at the center of their gain spectra, while LL-9.8 has a negative GVD around $-2000$ s$^2$/mm. The change in sign of the GVD is expected because the zero-GVD point of InP is around 5.5 $\mu$m.

Although the theoretical analysis presented in this work has for the most part neglected GVD, the purpose of this
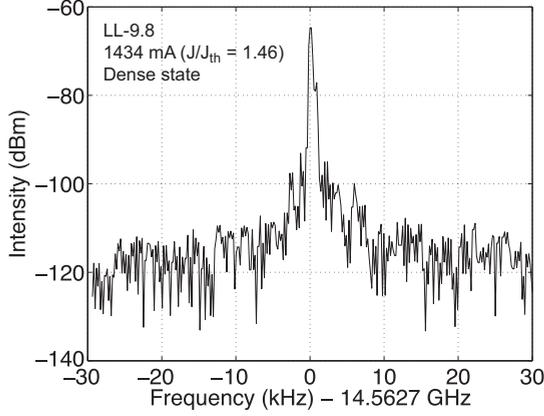
FIG. 10. The beat note of LL-9.8 at 1434 mA, the current at which the dense state appeared. The resolution bandwidth of the spectrum analyzer is 430 Hz.

measurement is to demonstrate that the harmonic state exists over a wide range of GVD, including both positive and negative values. We hope that including these data here will help guide future work on the role played by GVD in the properties of the harmonic state.

## APPENDIX C: BEAT NOTE IN DENSE STATE

In the dense state the mode spacing is one FSR ($\approx$15 GHz for a 3-mm-long QCL), which is low enough for the beat note to be measured by standard techniques, unlike the case of the harmonic state. We measure the beat note by the electrical technique: the current source is passed from the dc to the ac+dc port of a high-frequency bias tee (Tektronix PSPL5544) en route to the QCL, so that current modulation induced by any intracavity intensity modulation can be measured by a spectrum analyzer connected to the AC port of the bias tee. In Fig. 10, the beat note of LL-9.8 is shown at 1434 mA, the current at which the dense state appeared on this particular upward ramp up of the current. The width of the beat note is on the order of a few kHz. Devices TL-4.6 and DS-3.8 exhibited similar beat notes at the onset of the dense state (not plotted). As the current was increased further, different regimes could be observed such as the appearance of multiple closely spaced beat notes, or a broadening of the beat note. Because the focus of this work is on the harmonic state, these diverse behaviors of the dense state will be explored in future work.

## APPENDIX D: GAIN RECOVERY TIME VS UPPER-STATE LIFETIME

In a QCL, the upper-state lifetime $T_{up}$ tells us how long an electron sits in the upper state before making a nonradiative transition to the lower state. From here, it takes some additional time to travel through the injector region and tunnel into the upper level of the next stage. This additional amount of time is the bottleneck that determines the gain recovery time. The Maxwell-Bloch equations, by making the two-level approximation, cannot account for the full complexity of the QCL, and only provide us with one carrier relaxation time, which we have called $T_1$. This begs the following question:

Does $T_1$ represent the upper-state lifetime or the gain recovery time? The answer is that it depends on what you want to calculate. In the steady-state single-mode regime, we find that the output power and population inversion are functions of $T_1$ due to diffusion; here, we argue that $T_1$ should represent the upper-state lifetime $T_{up}$ because $T_{up}$ tells us how much time an electron in the upper state has to diffuse before transitioning to the lower state. It is for this reason that $T_{up}$ appears in the definition of $\gamma_D$, $\gamma_D = (1 + 4k^2 D T_{up})^{-1}$, rather than $T_1$. In dynamical situations, on the other hand, the intensity of the field varies with time and we are interested in how the population inversion responds. We argue that this response is determined by the gain recovery time, not the upper-state lifetime, because the response definitely depends on how long it takes an electron to get from one active stage to the next. To summarize it concisely, the upper-state lifetime is used for the calculation of the population grating (PG), but the gain recovery time is used for the calculation of the population pulsations (PPs). Since the bulk of our paper deals with PPs, we chose simply to call $T_1$ the gain recovery time, rather than name a new time scale such as $T_{gr}$, for instance. We hope that this does not confuse the reader.

## APPENDIX E: THEORY: SINGLE-MODE SOLUTION

This Appendix gives a more detailed derivation of the single-mode solution presented in Sec. III B, including the intracavity power as a function of pumping, and the population inversion as a function of position and pumping.

For a two-level system with upper state $|a\rangle$ and lower state $|b\rangle$, the material equations in the nonrotating frame and the field equation are

$$\frac{d\rho_{ab}}{dt} = -i\omega_{ba}\rho_{ab} - \frac{id}{\hbar}E(t)w - \frac{\rho_{ab}}{T_2}, \qquad (E1)$$

$$\frac{dw}{dt} = \frac{-2id}{\hbar}E(t)(\rho_{ab} - \rho_{ab}^*) + \frac{w_{eq} - w}{T_1} + D\frac{\partial^2 w}{\partial z^2}, \quad (E2)$$

$$\frac{\partial^2 E}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = Nd\mu\frac{\partial^2}{\partial t^2}(\rho_{ab} + \rho_{ab}^*). \qquad (E3)$$

We emphasize that these equations are in the nonrotating frame, whereas the equations we have used in the main text [51] were already in the rotating frame and the RWA had already been applied. However, since we are here dealing with two counterpropagating waves, we chose to more closely follow the approach in [33]. We make the following *Ansätze*:

$$E(z,t) = \frac{1}{\sqrt{2}}\big[\mathcal{E}_R(z,t)e^{-i(\omega t - kz)} + \mathcal{E}_L(z,t)e^{-i(\omega t + kz)} + \text{c.c.}\big],$$

$$\tag{E4}$$

$$\rho_{ab}(z,t) = \eta_R^*(z,t)e^{-i(\omega t - kz)} + \eta_L^*(z,t)e^{-i(\omega t + kz)}, \qquad (E5)$$

$$w(z,t) = w_{DC}(z,t) + w_2(z,t)e^{i2kz} + w_2^*(z,t)e^{-i2kz}. \quad (E6)$$

(We use the subscript "DC" rather than "0" for the spatial average of the population inversion, $w_{DC}$, because the subscript 0 is used throughout the text to refer to the primary mode. No such ambiguity occurs for the subscript "2.") Plugging the *Ansätze* into the differential equations, and making the RWA

as well as the slowly varying envelope approximation (SVEA) yields the following equations:

$$\frac{d\eta_R^*}{dt} = \frac{-i\kappa}{2\sqrt{2}}(\mathcal{E}_R w_{\text{DC}} + \mathcal{E}_L w_2) - \left(\frac{1}{T_2} + i\Delta\right)\eta_R^*, \quad \text{(E7)}$$

$$\frac{d\eta_L^*}{dt} = \frac{-i\kappa}{2\sqrt{2}}(\mathcal{E}_L w_{\text{DC}} + \mathcal{E}_R w_2^*) - \left(\frac{1}{T_2} + i\Delta\right)\eta_L^*, \quad \text{(E8)}$$

$$\frac{dw_{\text{DC}}}{dt} = \frac{i\kappa}{\sqrt{2}}(\mathcal{E}_R \eta_R + \mathcal{E}_L \eta_L - \text{c.c.}) + \frac{w_{\text{eq}} - w_{\text{DC}}}{T_1}, \quad \text{(E9)}$$

$$\frac{dw_2}{dt} = \frac{i\kappa}{\sqrt{2}}(\mathcal{E}_R \eta_L - \mathcal{E}_L^* \eta_R^*) - \frac{w_2}{T_1} - 4k^2 D w_2, \quad \text{(E10)}$$

$$\frac{1}{c}\frac{\partial \mathcal{E}_R}{\partial t} = -\frac{\partial \mathcal{E}_R}{\partial z} + \frac{i\sqrt{2}\alpha}{\kappa T_2}\eta_R^* - \frac{\ell_0}{2}\mathcal{E}_R, \quad \text{(E11)}$$

$$\frac{1}{c}\frac{\partial \mathcal{E}_L}{\partial t} = +\frac{\partial \mathcal{E}_L}{\partial z} + \frac{i\sqrt{2}\alpha}{\kappa T_2}\eta_L^* - \frac{\ell_0}{2}\mathcal{E}_L, \quad \text{(E12)}$$

where $\kappa = 2d/\hbar$, $\alpha = N\omega T_2 d^2\sqrt{\mu/\epsilon}/\hbar$ is the Beer coefficient of the material, and $\Delta = \omega_{\text{ba}} - \omega$ is the detuning of the field from the atomic resonance frequency. The loss term $\ell_0$ has been added to the field equation heuristically, and in this context it represents only the waveguide loss.

We solve for the single-mode solution by setting the time derivatives to zero and the slowly varying envelope functions to be constants. In doing so, we are now making the distributed loss approximation because we are not allowing the fields to grow in space. Thus, $\ell_0$ must now be taken to be the total loss, waveguide plus mirror loss, which we call $\ell$. (In the main text, the loss is expressed by the rate $\bar{\ell}$, which is simply given by $\bar{\ell} \equiv \ell c$, and the Beer coefficient $\alpha$ is similarly converted into a rate $\bar{\alpha} \equiv \alpha c$.) We take $\Delta = 0$ for simplicity because the single mode will lase very close to the peak of the gain spectrum. We denote the steady-state field amplitudes by $\mathcal{E}_R = \mathcal{E}_L = \mathcal{E}_0$ and find the LI curve

$$|\tilde{\mathcal{E}}_0|^2 = \frac{p-1}{1+\gamma_D/2}, \quad \text{(E13)}$$

where $\tilde{\mathcal{E}}_0 \equiv \kappa\sqrt{T_1 T_2}\mathcal{E}_0$, $p = w_{\text{eq}}/w_{\text{th}}$, $w_{\text{th}} = \bar{\ell}/\bar{\alpha}$, and $\gamma_D = (1 + 4k^2 D T_1)^{-1}$ is the diffusion parameter. Based on the discussion in Appendix D, however, we know $T_1$ represents the upper-state lifetime $T_{\text{up}}$, so we redefine $\gamma_D = (1 + 4k^2 D T_{\text{up}})^{-1}$. The steady-state population $w_0(z)$ is given by

$$w_0(z) = w_{\text{th}}\left[1 + \frac{\gamma_D}{2}\frac{p-1}{1+\gamma_D/2} - \gamma_D\frac{p-1}{1+\gamma_D/2}\cos(2k_0 z)\right]. \quad \text{(E14)}$$

As the diffusivity or $T_{\text{up}}$ increases, causing $\gamma_D$ to approach zero, the population grating is "washed out" and $w_0(z)$ is uniformly equal to $w_{\text{th}}$.

We have made the approximation that the pumping $p$ is uniform in space. For an electrically injected QCL, this is equivalent to assuming that the injected current density $J$ is uniform throughout the cavity. In fact, however, the resistance of the active region is lower in the field antinodes because stimulated emission increases the rate of electron transport. Therefore, assuming a constant voltage across the active region, more current will flow through the lower-resistance antinodes, an effect which reduces the amplitude of the population grating even in the absence of any lateral carrier

diffusion. It helps to picture the active region as two resistors in series, one constant "background" resistance in series with one whose resistance drops with increasing light intensity. Devices with a lower background resistance will be prone to greater current inhomogeneity. The magnitude of this effect can be estimated from the kink in the current-voltage curve above threshold, which shows how much the photon field reduces the device resistance, but we have ignored this effect in our current work. One consequence of ignoring this inhomogeneity of the current is that we overestimate the amplitude of the population grating. This likely contributes to our theory's underestimation of the instability threshold $p_{\text{sb}}$, as discussed in Appendix H.

## APPENDIX F: THEORY: POPULATION PULSATIONS

This Appendix gives a more detailed derivation of the population pulsations presented in Sec. III C, and demonstrates how to include nonzero detuning $\Delta$ and GVD into the formalism.

We begin by imagining a small volume of dipoles subject to a spatially uniform $E$ field to develop an understanding of the nonlinear effects caused by the Bloch dynamics. The electric field is given by

$$E(t) = \mathcal{E}(t)e^{i\omega t} + \text{c.c.} \quad \text{(F1)}$$

The Bloch equations in the rotating-wave approximation are

$$\dot{\sigma} = \left(i\Delta - \frac{1}{T_2}\right)\sigma + \frac{i\kappa}{2}w\mathcal{E}, \quad \text{(F2)}$$

$$\dot{w} = i\kappa(\mathcal{E}^*\sigma - \mathcal{E}\sigma^*) - \frac{w - w_{\text{eq}}}{T_1}, \quad \text{(F3)}$$

where $\sigma$ is the off-diagonal element of the density matrix in the rotating frame, $w$ is the population inversion (positive when inverted), $\Delta = \omega_{\text{ba}} - \omega$ is the detuning between the applied field and the resonant frequency of the two-level system, $T_1$ is the (longitudinal) population relaxation time, $T_2$ is the (transverse) dephasing time, $\kappa \equiv 2d/\hbar$ is the coupling constant where $d$ is the dipole matrix element (assumed to be real) and $\hbar$ is Planck's constant, and $w_{\text{eq}}$ is the equilibrium population inversion in the absence of any electric field which is determined by the pumping. [Note that these equations are identical to Eqs. 3.19(a)–3.19(c) in [51], except that we have allowed $\mathcal{E}$ to be complex and left the off-diagonal component of the density matrix in complex notation rather than writing $\sigma = (u + iv)/2$.] With these conventions, the macroscopic polarization $P$ (dipole moment per volume) in a region with a volume density of $N$ dipoles is given by

$$P(t) = Nd\sigma e^{i\omega t} + \text{c.c.} \quad \text{(F4)}$$

First, we consider the effect of a monochromatic field at frequency $\omega$, obtained from Eqs. (F2) and (F3) by setting $\mathcal{E}(t) = \mathcal{E}_0$ and all time derivatives to zero. The result is a steady-state polarization $\sigma_0$ and population inversion $w_0$ given by

$$\sigma_0 = \frac{i\kappa T_2}{2(1 - i\Delta T_2)}w_0\mathcal{E}_0, \quad \text{(F5)}$$

$$w_0 = \frac{w_{\text{eq}}}{1 + \frac{\kappa^2 T_1 T_2 |\mathcal{E}_0|^2}{1+(\Delta T_2)^2}}. \quad \text{(F6)}$$

Note that the population inversion $w_0$ is saturated as the field strength $\mathcal{E}_0$ increases: this is responsible for saturable loss (when $w_{\text{eq}} < 0$) and saturable gain (when $w_{\text{eq}} > 0$).

N.B. In our equations so far, we have said the frequency of the field is $\omega$. Later on, we refer to the primary mode frequency as $\omega_0$. For our purposes here, $\omega$ and $\omega_0$ are interchangeable. In future work, this will not be the case. In analogy with the theory developed for microresonators [65], we plan in future work to adopt the convention that $\omega_0$ represents the center frequency of the cold-cavity mode that is lasing. However, the lasing frequency $\omega$ can be detuned from this cold-cavity resonance due to small frequency-pulling effects when one accounts for the hot cavity. This detuning is an important parameter in microresonators, where the pump frequency $\omega$ can be controllably tuned away from $\omega_0$, allowing one to compensate for GVD and optimize comb generation. We have not accounted for such a detuning in our work, and therefore $\omega$ and $\omega_0$ are interchangeable.

### 1. Two-frequency operation

Next, we consider the $E$ field

$$\mathcal{E} = \mathcal{E}_0 + \mathcal{E}_+ e^{i\delta\omega t} \tag{F7}$$

which consists of the strong field $\mathcal{E}_0$ at frequency $\omega$ superposed with the much weaker field $\mathcal{E}_+$ detuned from $\omega$ by $\delta\omega$. A polarization will of course be induced at $\omega + \delta\omega$. However, a polarization at $\omega - \delta\omega$ also results due to the beat note at $\delta\omega$ which modulates the intensity: the resulting modulation of the population inversion with time (i.e., a population pulsation) leads to nonlinear frequency mixing. We express the full polarization as

$$P(t) = \sum_{m=-,0,+} \mathcal{P}_m e^{i\omega_m t} + \text{c.c.}, \tag{F8}$$

where $\omega_+ \equiv \omega + \delta\omega$ and $\omega_- \equiv \omega - \delta\omega$. We can solve for the polarization as done in [58], keeping only terms to first order in the weak field $\mathcal{E}_+$, which gives

$$\mathcal{P}_0 = \frac{i\epsilon}{\omega_{\text{ba}}} \bar{\alpha} w_0 \frac{\mathcal{E}_0}{1 - i\Delta T_2}, \tag{F9}$$

$$\mathcal{P}_+ = \frac{i\epsilon}{\omega_{\text{ba}}} \bar{\alpha} w_0 \left[ \frac{\mathcal{E}_+}{1 - i(\Delta - \delta\omega)T_2} + \Lambda_+^+ \tilde{\mathcal{E}}_0 \tilde{\mathcal{E}}_0^* \mathcal{E}_+ \right], \tag{F10}$$

$$\mathcal{P}_- = \frac{i\epsilon}{\omega_{\text{ba}}} \bar{\alpha} w_0 \Lambda_-^+ \tilde{\mathcal{E}}_0 \tilde{\mathcal{E}}_0 \mathcal{E}_+^*, \tag{F11}$$

where

$$\Lambda_+^+ = \frac{-(1 + i\delta\omega T_2/2)[1 + i(\Delta + \delta\omega)T_2]/[1 - i(\Delta - \delta\omega)T_2]}{(1 + i\Delta T_2)\{(1 + i\delta\omega T_1)[1 + i(\Delta + \delta\omega)T_2][1 - i(\Delta - \delta\omega)T_2] + (1 + i\delta\omega T_2)|\tilde{\mathcal{E}}_0|^2\}}, \tag{F12}$$

$$\Lambda_-^+ = \frac{-(1 - i\delta\omega T_2/2)}{(1 - i\Delta T_2)\{(1 - i\delta\omega T_1)[1 + i(\Delta - \delta\omega)T_2][1 - i(\Delta + \delta\omega)T_2] + (1 - i\delta\omega T_2)|\tilde{\mathcal{E}}_0|^2\}} \tag{F13}$$

are the self-mixing and cross-mixing coupling coefficients, respectively. We consider the dipoles to be embedded in a host medium of permittivity $\epsilon$ and permeability $\mu$. (We adopt the convention of [52]: $\epsilon$, $\mu$ and the speed of light $c = 1/\sqrt{\epsilon\mu}$ always take their values in the background host medium.) Many of the material properties of the two-level system are lumped into the "Beer rate"

$$\bar{\alpha} = \frac{Nd^2 T_2 \omega_{\text{ba}} c\sqrt{\mu/\epsilon}}{\hbar}, \tag{F14}$$

which is related to the more familiar Beer absorption coefficient $\alpha$ (with units of inverse length) that appears in Beer's law of absorption by $\bar{\alpha} = \alpha c$. (Note, however, that in our expressions for the polarization due to the two-level system, all factors of $\epsilon$ and $\mu$ drop out; that is, these expressions do not contain the polarization contributions due to the background medium.) The central mode amplitude $\mathcal{E}_0$ has been normalized such that $\tilde{\mathcal{E}}_0 \equiv \kappa \sqrt{T_1 T_2} \mathcal{E}_0$. Note that $\mathcal{P}_0$ is unaffected to first order in $\mathcal{E}_+$. The polarization $\mathcal{P}_+$ comes from two contributions. First, there is the linear contribution from the Lorentz oscillator which $\mathcal{E}_+$ would induce even in the absence of the strong field $\mathcal{E}_0$. Second, there is a contribution due to the PP which is described by the term $\Lambda_+^+$. The term $\mathcal{P}_-$ is due solely to the PP and is governed by $\Lambda_-^+$. Note that the full polarization is directly proportional to the steady-state population inversion $w_0$; this will be important when we generalize our results to standing-wave cavities, where $w_0$ varies with position.

Now that we have the polarization, we can calculate the gain seen by the sideband field. We define the gain $\bar{g}$ (with dimension of frequency) of the sideband as the power density generated at $\omega + \delta\omega$ by the interaction of the field with the dipoles, considering only field and polarization terms oscillating at $\omega + \delta\omega$, divided by the energy density of the exciting sideband field, or

$$\bar{g}_+ \equiv -\frac{\langle E\dot{P}\rangle_+}{2\epsilon|\mathcal{E}_+|^2} \tag{F15}$$

$$= \frac{i\omega_+(\mathcal{E}_+\mathcal{P}_+^* - \mathcal{E}_+^*\mathcal{P}_+)}{2\epsilon|\mathcal{E}_+|^2}. \tag{F16}$$

#### a. $\Delta = 0$

Here, we consider the case of zero detuning $\Delta = 0$, which simplifies the mathematical expressions considerably. Under this scenario, we denote the self-mixing coefficient $\Lambda_+^+$ by $\Lambda$, where

$$\Lambda = \frac{-(1 + i\delta\omega T_2/2)}{[(1 + i\delta\omega T_1)(1 + i\delta\omega T_2)^2 + (1 + i\delta\omega T_2)|\tilde{\mathcal{E}}_0|^2]}, \tag{F17}$$

and it is simple to show that the cross-coupling coefficient $\Lambda_-^+$ is $\Lambda^*$. The gain of the sideband field is found to be

$$\bar{g}_+ = \bar{\alpha} w_0 \left[ \frac{1}{1 + (\delta\omega T_2)^2} + \text{Re}(\Lambda)|\tilde{\mathcal{E}}_0|^2 \right]. \tag{F18}$$

[We have used $(\omega + \delta\omega)/\omega_{ba} \approx 1$.] Thus, the gain can be nicely divided up into a contribution from the Lorentz oscillator and a contribution from the PP. All of this is proportional to $\bar\alpha w_0$: $\bar\alpha$ gives you the gain of a weak field tuned to line center in a perfectly inverted medium (or alternatively, the loss seen by a weak field tuned to line center in a material in its ground state), and $w_0$ gives you the expectation value of finding an electron in the excited state (equal to 1 when excited, $-1$ when in the ground state, and 0 at transparency). Note that $\mathrm{Re}(\Lambda)$ can be positive or negative, which we will discuss shortly.

### 2. Three-frequency operation

Of course, the polarization created at $\omega - \delta\omega$ will create a field at that frequency, which is precisely why in the experiments we always observe the two sidebands appearing simultaneously. One sideband cannot exist in isolation when the mixing terms naturally couple them together. Therefore, we need to consider the field

$$\mathcal{E} = \mathcal{E}_0 + \mathcal{E}_+ e^{i\delta\omega t} + \mathcal{E}_- e^{-i\delta\omega t}. \quad (F19)$$

The polarization at each sideband frequency now contains a Lorentzian term, a self-mixing term, and a cross-mixing term:

$$\mathcal{P}_+ = \frac{i\epsilon}{\omega_{ba}} \bar\alpha w_0$$
$$\times \left\{ \frac{\mathcal{E}_+}{[1 - i(\Delta - \delta\omega)T_2]} + \Lambda_+^+ \tilde{\mathcal{E}}_0 \tilde{\mathcal{E}}_0^* \mathcal{E}_+ + \Lambda_+^- \tilde{\mathcal{E}}_0 \tilde{\mathcal{E}}_0 \mathcal{E}_-^* \right\}, \quad (F20)$$

$$\mathcal{P}_- = \frac{i\epsilon}{\omega_{ba}} \bar\alpha w_0$$
$$\times \left\{ \frac{\mathcal{E}_-}{[1 - i(\Delta + \delta\omega)T_2]} + \Lambda_-^- \tilde{\mathcal{E}}_0 \tilde{\mathcal{E}}_0^* \mathcal{E}_- + \Lambda_-^+ \tilde{\mathcal{E}}_0 \tilde{\mathcal{E}}_0 \mathcal{E}_+^* \right\}, \quad (F21)$$

where $\Lambda_-^-$ and $\Lambda_+^-$ are obtained by making the substitution $\delta\omega \to -\delta\omega$ in the expressions for $\Lambda_+^+$ and $\Lambda_-^+$, respectively, given in Eqs. (F12) and (F13).

### *a.* $\Delta = 0$

Let us again focus on the case $\Delta = 0$, for which the polarization at each sideband simplifies to

$$\mathcal{P}_+ = \frac{i\epsilon}{\omega_{ba}} \bar\alpha w_0 \left\{ \frac{\mathcal{E}_+}{[1 + i\delta\omega T_2]} + \Lambda \tilde{\mathcal{E}}_0 \tilde{\mathcal{E}}_0^* \mathcal{E}_+ + \Lambda \tilde{\mathcal{E}}_0 \tilde{\mathcal{E}}_0 \mathcal{E}_-^* \right\}, \quad (F22)$$

$$\mathcal{P}_- = \frac{i\epsilon}{\omega_{ba}} \bar\alpha w_0 \left\{ \frac{\mathcal{E}_-}{[1 - i\delta\omega T_2]} + \Lambda^* \tilde{\mathcal{E}}_0 \tilde{\mathcal{E}}_0^* \mathcal{E}_- + \Lambda^* \tilde{\mathcal{E}}_0 \tilde{\mathcal{E}}_0 \mathcal{E}_+^* \right\}, \quad (F23)$$

where $\Lambda$ is simply $\Lambda_+^+$ evaluated for $\Delta = 0$. We see the nice property that when $\Delta = 0$, $\Lambda_+^+ = \Lambda_+^-$ ($\equiv \Lambda$), and $\Lambda_-^- = \Lambda_-^+$ ($\equiv \Lambda^*$); in other words, the self- and cross-mixing coupling coefficients are equal.

The gain $\bar{g}_+$ of the positive sideband is

$$\bar{g}_+ = \bar\alpha w_0 \left\{ \frac{1}{1 + (\delta\omega T_2)^2} + \mathrm{Re}\left[ \Lambda |\tilde{\mathcal{E}}_0|^2 \left( 1 + \frac{\tilde{\mathcal{E}}_0^2 \mathcal{E}_-^*}{|\tilde{\mathcal{E}}_0|^2 \mathcal{E}_+} \right) \right] \right\}, \quad (F24)$$

and a similar expression holds for the minus sideband. This equation tells us that the PP contribution to the gain depends on the phase and amplitude relationships of $\mathcal{E}_0$, $\mathcal{E}_-$, and $\mathcal{E}_+$, which is not too surprising because the amplitude of the PP itself is sensitive to these parameters. Without loss of generality, we can take $\mathcal{E}_0$ to be real. If $\mathcal{E}_+ = \mathcal{E}_-^*$, then the two sidebands' contributions to the beat note at $\delta\omega$ add constructively, resulting in a field whose amplitude modulation (AM) is twice the strength of a field with only one sideband. If $\mathcal{E}_+ = -\mathcal{E}_-^*$, then the two sidebands' contributions to the beat note at $\delta\omega$ destructively cancel and there is no longer any amplitude modulation at frequency $\delta\omega$. We refer to such a field as frequency modulated (FM). We see from Eq. (F24) that the AM sidebands therefore experience a PP contribution to the gain that is twice as large as the single sideband case, while the FM sidebands experience only the background Lorentzian gain, consistent with the fact that there is no PP in this case. We summarize this with the formula for the gain $\bar{g}$ of each sideband for the case of equal-amplitude sidebands ($|\mathcal{E}_+| = |\mathcal{E}_-|$)

$$\bar{g} = \bar\alpha w_0 \left[ \frac{1}{1 + (\delta\omega T_2)^2} + \mathrm{Re}(\Lambda)|\tilde{\mathcal{E}}_0|^2 \times \begin{cases} 2 & (\mathrm{AM}) \\ 0 & (\mathrm{FM}) \end{cases} \right]. \quad (F25)$$

Note that for a superposition of AM and FM, the gain due to the PP will fall between 0 and 2 times the factor $\mathrm{Re}(\Lambda)|\tilde{\mathcal{E}}_0|^2$.

### APPENDIX G: THEORY: INSTABILITY THRESHOLD

This Appendix gives a more detailed derivation of the instability threshold presented in Sec. III D, and demonstrates that the gain seen by the sidebands is due to a contribution from the population grating and another from the population pulsations.

When a continuous-wave (cw) laser is pumped at its lasing threshold, only a single frequency of light, the one nearest the gain peak that also satisfies the round-trip phase condition, has sufficient gain to overcome the round-trip loss and begins to lase. As the pumping is increased, the single-mode solution yields to multimode operation; this is known as the single-mode instability. Our goal is to determine (1) how hard to pump the laser to reach the single-mode instability and (2) which new frequencies start lasing.

Consider a laser pumped above threshold that is lasing on a single mode, which we refer to as the primary or central mode. If another mode is to lase, it must be seeded by a spontaneously generated photon at a different frequency. This photon will necessarily create a beat note through its coexistence with the primary mode, resulting in a population pulsation. The gain seen by the new frequency must therefore account for this parametric gain in addition to the background Lorentzian gain. Furthermore, the PP couples the sideband to the symmetrically detuned sideband frequency on the other side of the primary

mode, so we should in general assume the presence of both sidebands. Because the instability threshold depends on the cavity geometry, we will consider a traveling-wave laser as well as a standing-wave laser. In both cases, the strategy is the same. First, we solve for the single-mode intensity $\mathcal{E}_0$ and the population inversion $w_0(z)$ as a function of the pumping, entirely neglecting the sidebands. Knowing this, we can then calculate the sideband gain in the presence of the primary mode.

We start with the wave equation

$$\frac{\partial^2 E}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = \mu \frac{\partial^2 P}{\partial t^2}. \tag{G1}$$

Following the approach used to calculate the optical parametric oscillation threshold in optically pumped microresonators [65], we expand the field in terms of the cold-cavity modes

$$E(z,t) = \sum_{m=-,0,+} \mathcal{E}_m(t)\Upsilon_m(z)e^{i\omega_m t} + \text{c.c.} \tag{G2}$$

The spatial modes obey the normalization condition

$$\frac{1}{L}\int_0^L dz\, |\Upsilon_m(z)|^2 = 1. \tag{G3}$$

When group velocity dispersion (GVD) is nonzero, the two modes $\omega_+$ and $\omega_-$ will not be equidistant from $\omega_0$. We have also assumed that the spatial and temporal dependence of the modes can be separated. This is a good approximation in the case of a laser because we know the intracavity field will be sharply resonant at the modes. The spatial variation of the polarization can be described by making the substitution $\mathcal{E}_m \to \mathcal{E}_m \Upsilon_m(z)$ and $w_0 \to w_0(z)$ into the polarization equations (F20) and (F21), which results in the polarization

$$P(z,t) = \sum_{m=-,0,+} \mathcal{P}_m(z,t)e^{i\omega_m t} + \text{c.c.}, \tag{G4}$$

where

$$\mathcal{P}_+(z,t) = \frac{i\epsilon}{\omega_{\text{ba}}}\bar{\alpha}w_0(z)\left\{\frac{\mathcal{E}_+\Upsilon_+(z)}{[1-i(\Delta-\delta\omega)T_2]} + \Lambda_+^+(z)|\Upsilon_0(z)|^2\Upsilon_+(z)|\tilde{\mathcal{E}}_0|^2\mathcal{E}_+ + \Lambda_+^-(z)\Upsilon_0(z)^2\Upsilon_-^*(z)e^{i\bar{\omega}t}\tilde{\mathcal{E}}_0^2\mathcal{E}_-^*\right\}, \tag{G5}$$

$$\mathcal{P}_-(z,t) = \frac{i\epsilon}{\omega_{\text{ba}}}\bar{\alpha}w_0(z)\left\{\frac{\mathcal{E}_-\Upsilon_-(z)}{[1-i(\Delta+\delta\omega)T_2]} + \Lambda_-^-(z)|\Upsilon_0(z)|^2\Upsilon_-(z)|\tilde{\mathcal{E}}_0|^2\mathcal{E}_- + \Lambda_-^+(z)\Upsilon_0(z)^2\Upsilon_+^*(z)e^{i\bar{\omega}t}\tilde{\mathcal{E}}_0^2\mathcal{E}_+^*\right\}. \tag{G6}$$

We have introduced $\bar{\omega} \equiv 2\omega_0 - \omega_+ - \omega_-$, the deviation of the cold-cavity modes from equal spacing. Note that the $\Lambda$'s now depend on $z$ due to the term in their denominators dependent on the primary mode amplitude. Because we no longer demand that the two sidebands have the same detuning $\delta\omega$, $\Lambda_+^+$ and $\Lambda_-^+$ should, strictly speaking, be calculated using the detuning $\delta\omega_+ = \omega_+ - \omega_0$, while $\Lambda_-^-$ and $\Lambda_+^-$ should depend on $\delta\omega_- = \omega_0 - \omega_-$. In practice, we can ignore this difference in the $\Lambda$'s; the term $e^{i\bar{\omega}t}$ captures the most important effect of GVD.

Plugging everything into the wave equation gives

$$\sum_m \left(\frac{d^2\Upsilon_m}{dz^2} + \frac{\omega_m^2}{c^2}\Upsilon_m\right)\mathcal{E}_m e^{i\omega_m t} - \frac{2i}{c^2}\sum_m \omega_m \frac{d\mathcal{E}_m}{dt}\Upsilon_m e^{i\omega_m t} = \mu\sum_m -\omega_m^2(\mathcal{P}_m - \mathcal{P}_{m,\text{loss}})e^{i\omega_m t}, \tag{G7}$$

where the slowly varying envelope approximation allowed us to ignore second time derivatives of $\mathcal{E}_m$ on the left-hand side, and first and second derivatives of $\mathcal{E}_m$ on the right-hand side. The spatial modes $\Upsilon_m(z)$ are chosen so that the first term on the left-hand side equals zero. The loss of each mode has been added to the equation in the form of a polarization contribution; we assume each mode has the same linear loss, which can be expressed

$$\mathcal{P}_{m,\text{loss}}(z,t) = \frac{i\epsilon}{\omega_{\text{ba}}}\bar{\ell}\Upsilon_m(z)\mathcal{E}_m(t). \tag{G8}$$

Equation (G7) couples all of the modes $\mathcal{E}_m$. We can project this equation onto each mode by multiplying by $\Upsilon_n(z)$ and integrating over the length of the laser cavity, thus taking advantage of the orthonormality of the spatial modes $\Upsilon_m(z)$, and then equating terms which oscillate at the same frequency (since terms with different frequencies will not affect the time-averaged gain seen by a mode). The result is one equation for the central mode

$$\dot{\mathcal{E}}_0 = \left[-\frac{\bar{\ell}}{2} + \frac{\bar{\alpha}}{2(1-i\Delta T_2)}\int\frac{dz}{L}\,w_0(z)|\Upsilon_0(z)|^2\right]\mathcal{E}_0, \tag{G9}$$

and one equation for each of the sidebands

$$\dot{\mathcal{E}}_+ = -\frac{\bar{\ell}}{2}\mathcal{E}_+ + \frac{\bar{\alpha}}{2}\left[\frac{\mathcal{E}_+}{1-i(\Delta-\delta\omega)T_2}\int\frac{dz}{L}\,w_0(z)|\Upsilon_+(z)|^2 + |\tilde{\mathcal{E}}_0|^2\mathcal{E}_+\int\frac{dz}{L}\,w_0(z)\Lambda_+^+(z)|\Upsilon_0(z)|^2|\Upsilon_+(z)|^2\right.$$
$$\left. + \tilde{\mathcal{E}}_0^2\mathcal{E}_-^*e^{i\bar{\omega}t}\int\frac{dz}{L}\,w_0(z)\Lambda_+^-(z)\Upsilon_0(z)^2\Upsilon_-^*(z)\Upsilon_+^*(z)\right], \tag{G10}$$

$$\dot{\mathcal{E}}_- = -\frac{\bar{\ell}}{2}\mathcal{E}_- + \frac{\bar{\alpha}}{2}\left[\frac{\mathcal{E}_+}{1-i(\Delta+\delta\omega)T_2}\int\frac{dz}{L}\,w_0(z)|\Upsilon_-(z)|^2 + |\tilde{\mathcal{E}}_0|^2\mathcal{E}_-\int\frac{dz}{L}\,w_0(z)\Lambda_-^-(z)|\Upsilon_0(z)|^2|\Upsilon_-(z)|^2\right.$$
$$\left. + \tilde{\mathcal{E}}_0^2\mathcal{E}_+^*e^{i\bar{\omega}t}\int\frac{dz}{L}\,w_0(z)\Lambda_-^+(z)\Upsilon_0(z)^2\Upsilon_+^*(z)\Upsilon_-^*(z)\right]. \tag{G11}$$

These three equations will be used to understand the instability threshold. In general, one must first apply the steady-state condition $\dot{\mathcal{E}}_0 = 0$ to Eq. (G9) which, together with the Bloch equation relating the field to the inversion, will yield the amplitude of the primary mode $\mathcal{E}_0$ along with the resulting population inversion $w_0(z)$, both as a function of the pumping $w_{\text{eq}}$. (The result will be the same as what we calculated for the single-mode solution in Appendix E.) This information is then used in Eqs. (G10) and (G11) to determine the minimum level of pumping $w_{\text{eq}}$ at which a pair of sidebands with detuning $\delta\omega$ experiences more gain than loss. This is the instability threshold.

So far, we have kept Eqs. (G9)–(G11) as general as possible to account for arbitrary spatial profiles, GVD, and detuning $\Delta$ between the lasing mode and the peak of the gain spectrum. From here on we will simplify the problem by taking $\bar{\omega} = 0$ (zero GVD) and $\Delta = 0$, and apply these conditions to the simplest possible traveling-wave and standing-wave cavities.

### 1. Traveling-wave cavity

For the traveling-wave laser, the spatial modes are

$$\Upsilon_m(z) = e^{-ik_m z} \tag{G12}$$

so every point in the cavity sees the same intensity. At and above threshold, the population inversion is everywhere saturated to the threshold inversion, so $w_0$ is independent of $z$. For $\Delta = 0$, the inversion is

$$w_0 = w_{\text{th}} \equiv \frac{\bar{\ell}}{\bar{\alpha}} \tag{G13}$$

and the intensity of the primary mode is given by

$$|\tilde{\mathcal{E}}_0|^2 = p - 1, \tag{G14}$$

where we have made use of the normalized primary mode amplitude $\tilde{\mathcal{E}}_0 \equiv \kappa\sqrt{T_1 T_2}\mathcal{E}_0$, and $p$ is the pumping parameter defined as $p \equiv w_{\text{eq}}/w_{\text{th}}$. Because $|\tilde{\mathcal{E}}_0|^2$ is independent of $z$, all of the $\Lambda$'s are independent of $z$. Furthermore, since both $w_0$ and the $\Lambda$'s are independent of $z$, they can be pulled out of the spatial integrals in Eqs. (G10) and (G11). These integrals are then equal to one, where we have used the zero GVD condition $\bar{\omega} = 0$ in order for the cross-overlap integral (the last integral in each equation) to equal one. The sideband equations become

$$\dot{\mathcal{E}}_+ = -\frac{\bar{\ell}}{2}\mathcal{E}_+ + \frac{\bar{\alpha}w_{\text{th}}}{2}\left[\frac{\mathcal{E}_+}{1 + i\delta\omega T_2} + \Lambda|\tilde{\mathcal{E}}_0|^2\mathcal{E}_+ + \Lambda\tilde{\mathcal{E}}_0^2\mathcal{E}_-^*\right], \tag{G15}$$

$$\dot{\mathcal{E}}_- = -\frac{\bar{\ell}}{2}\mathcal{E}_- + \frac{\bar{\alpha}w_{\text{th}}}{2}\left[\frac{\mathcal{E}_-}{1 - i\delta\omega T_2} + \Lambda^*|\tilde{\mathcal{E}}_0|^2\mathcal{E}_- + \Lambda^*\tilde{\mathcal{E}}_0^2\mathcal{E}_+^*\right], \tag{G16}$$

which can be written in matrix form

$$\begin{pmatrix} \dot{\mathcal{E}}_+ \\ \dot{\mathcal{E}}_-^* \end{pmatrix} = \begin{pmatrix} M_+ & R_+ \\ R_-^* & M_-^* \end{pmatrix}\begin{pmatrix} \mathcal{E}_+ \\ \mathcal{E}_-^* \end{pmatrix}, \tag{G17}$$

where

$$M_+ = M_-^* = -\frac{\bar{\ell}}{2} + \frac{\bar{\alpha}w_{\text{th}}}{2}\left(\frac{1}{1 + i\delta\omega T_2} + \Lambda|\tilde{\mathcal{E}}_0|^2\right), \tag{G18}$$

$$R_+ = R_-^* = \frac{\bar{\alpha}w_{\text{th}}}{2}\Lambda|\tilde{\mathcal{E}}_0|^2. \tag{G19}$$

(In the last step, we have finally taken the freedom to choose $\tilde{\mathcal{E}}_0$ to be real, which we can do at this point without loss of generality.)

Now, if we assume a solution of the form $\mathcal{E}_\pm \sim e^{\lambda t}$, we find the two solutions for $\lambda$

$$\lambda = \tfrac{1}{2}[M_+ + M_-^* \pm \sqrt{(M_+ - M_-^*)^2 + 4R_+ R_-^*}]. \tag{G20}$$

The net gain seen by each sideband is given by $\text{Re}(2\lambda)$ (the factor of 2 is for intensity gain rather than amplitude gain), which includes the gain minus the loss. Subtracting off the loss, the gain $\bar{g}$ seen by each sideband is

$$\bar{g} = \bar{\alpha}w_{\text{th}}\left[\frac{1}{1 + (\delta\omega T_2)^2} + \begin{cases} 2\,\text{Re}(\Lambda)|\tilde{\mathcal{E}}_0|^2 & \text{(AM)} \\ 0 & \text{(FM)} \end{cases}\right], \tag{G21}$$

where the two solutions correspond to AM and FM sideband configurations. Finally, we recognize that the gain is pinned at threshold, so $\bar{\alpha}w_{\text{th}} = \bar{\ell}$, and we write the sideband gain normalized to the loss

$$\frac{\bar{g}}{\bar{\ell}} = \frac{1}{1 + (\delta\omega T_2)^2} + \text{Re}(\Lambda)|\tilde{\mathcal{E}}_0|^2 \times \begin{cases} 2 & \text{(AM)} \\ 0 & \text{(FM)} \end{cases}. \tag{G22}$$

When the gain $\bar{g}$ exceeds the loss $\bar{\ell}$, the weak sideband amplitudes experience exponential growth, therefore the single-mode solution becomes unstable. Note that the Lorentzian term is always less than 1. This is a direct result of uniform gain clamping in the traveling-wave laser, which clamps the net gain of the mode at the peak of the Lorentzian to zero, and therefore any mode detuned from the peak will see slightly more loss than gain. FM sidebands, therefore, never become unstable because they only see the Lorentzian gain. On the other hand, AM sidebands induce a PP and with it a coherent gain term, which can provide enough extra gain on top of the Lorentzian background to allow the sidebands to lase:

$$\frac{\bar{g}_{\text{AM}}}{\bar{\ell}} = \frac{1}{1 + (\delta\omega T_2)^2} + 2\,\text{Re}(\Lambda)|\tilde{\mathcal{E}}_0|^2. \tag{G23}$$

To get a feel for the sideband gain, we have plotted $\bar{g}_{\text{AM}}/\bar{\ell}$ in Fig. 11 at various pump strengths $p$ for $Z = 1, 10$, and $100$, where $Z \equiv T_1/T_2$. Graphically, we see that at large enough $p$, sidebands will become unstable. Analytically, it is a simple matter to calculate how hard to pump the laser $p$ before the sidebands appear, starting from Eq. (G23). We start by replacing $|\tilde{\mathcal{E}}_0|^2$ with $p - 1$, and note that this substitution must also be made in $\Lambda$, which implicitly varies with $|\tilde{\mathcal{E}}_0|^2$. Then, setting $\bar{g}_{\text{AM}}/\bar{\ell}$ equal to one, we can solve a simple quadratic formula for $\delta\omega^2$:

$$(\delta\omega T_2)^2$$
$$= \frac{-1 + 3Z(p-1) \pm \sqrt{[1 - 3Z(p-1)]^2 - 8Z^2 p(p-1)}}{2Z^2}. \tag{G24}$$

Finally, we must apply some physical reasoning: as $p$ is increased past 1, the sideband gain increases. Right at the moment when the instability threshold is reached, $\delta\omega^2$ must take on a single value. Thus, we set the radical in Eq. (G24) to zero and solve for $p$. After solving another simple quadratic
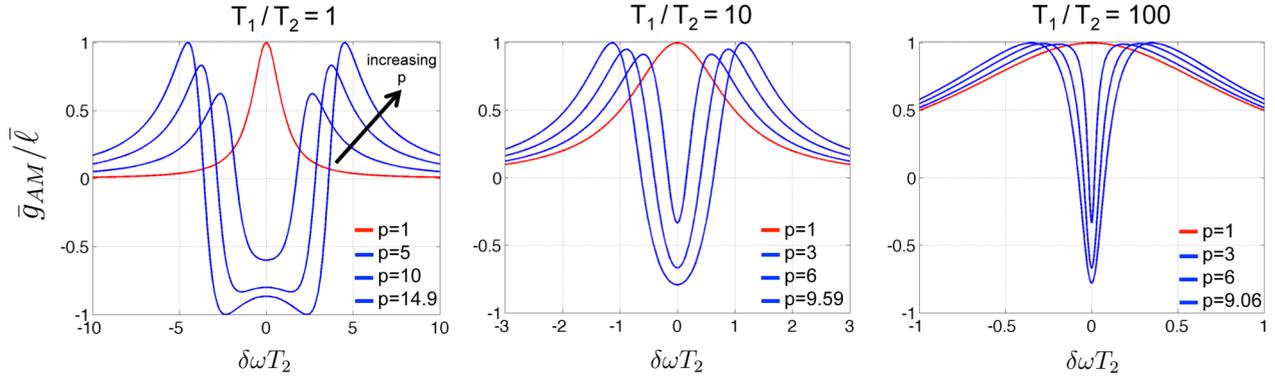
FIG. 11. The sideband gain $\bar{g}_{AM}/\bar{\ell}$ of a traveling-wave laser, given in Eq. (G23), is plotted at various pump strengths, for three different values of $T_1/T_2$: 1, 10, and 100. The largest value of $p$ in each plot is equal to the instability threshold given in Eq. (G26).

equation, we find that

$$p = 5 + \frac{3}{Z} \pm 4\sqrt{1 + \frac{3}{2Z} + \frac{1}{2Z^2}}. \quad (G25)$$

How do we choose between the plus and minus sign? By plugging this expression for $p$ back into Eq. (G24), it is simple to check that only the plus sign yields real-valued solutions for $\delta\omega$. Thus, we have found the instability threshold, which we denote $p_{\mathrm{RNGH}}$,

$$p_{\mathrm{RNGH}} = 5 + \frac{3}{Z} + 4\sqrt{1 + \frac{3}{2Z} + \frac{1}{2Z^2}} \quad (G26)$$

because it is the well-known instability threshold found by Risken and Nummedal [see Eq. (3.10) in [10]] and Graham and Haken [see Eq. (7.35) in [11]]. Plugging this value of $p$ into Eq. (G24) yields the value of $\delta\omega$ of the sidebands when the instability sets in:

$$(\delta\omega_{\mathrm{RNGH}}T_2)^2 = \frac{4}{Z^2} + \frac{6}{Z}\left(1 + \sqrt{1 + \frac{3}{2Z} + \frac{1}{2Z^2}}\right). \quad (G27)$$

One thing to notice is that in the limit $Z \gg 1$ (transverse relaxation must be faster than longitudinal relaxation), the instability threshold $p_{\mathrm{RNGH}} \to 9$ from above and $\delta\omega_{\mathrm{RNGH}}T_2 \to \sqrt{12/Z}$.

### 2. Standing-wave cavity

As before, we restrict ourselves to the case $\Delta = 0$ and $\bar{\omega} = 0$. We will see that calculations for the standing-wave cavity are significantly more complicated than for the traveling-wave cavity. The spatial variation of the primary mode causes the inversion $w_0$ and the coupling $\Lambda$ to both depend on $z$, which makes the integrals more difficult to compute. For this reason, we treat the problem to first order in the primary mode intensity $|\tilde{\mathcal{E}}_0|^2$, which allows us to compute the integrals analytically. However, the theory can be extended to higher order at will, or the integrals can always be computed numerically.

For the standing-wave laser with perfectly reflecting end mirrors, the spatial profile of each mode is given by

$$\Upsilon_m(z) = \sqrt{2}\cos(k_m z). \quad (G28)$$

The spatial modulation of the intensity is responsible for the spatial modulation of the population inversion $w_0(z)$,

given by Eq. (E14). This population grating has important consequences. For one, it reduces the power of the laser, which is given by Eq. (E13). Second, the gain is no longer uniformly clamped by the primary lasing mode, which will allow new modes to lase even in the absence of PPs.

The spatial variation of the primary lasing mode also causes $\Lambda$ to vary with position. In keeping with our approximations, we can expand $\Lambda$ to zeroth order in $|\tilde{\mathcal{E}}_0|^2$ because in our equations $\Lambda$ always multiplies $|\tilde{\mathcal{E}}_0|^2$, so the final result is first order in $|\tilde{\mathcal{E}}_0|^2$. We define the zeroth-order expansion of $\Lambda$ to be

$$\chi^{(3)} = \frac{-(1 + i\delta\omega T_2/2)}{(1 + i\delta\omega T_1)(1 + i\delta\omega T_2)^2}, \quad (G29)$$

where the symbol $\chi^{(3)}$ was chosen to emphasize that this term now plays the role of a third-order nonlinear coefficient.

We start with the sideband equations (G10) and (G11), replace $w_0(z)$ with Eq. (E14), $\Lambda(z)$ with $\chi^{(3)}$, and keep only terms to first order in $|\tilde{\mathcal{E}}_0|^2$. The resulting equation for the growth of the positive sideband is

$$\dot{\mathcal{E}}_+ = -\frac{\bar{\ell}}{2}\mathcal{E}_+ + \frac{\bar{\alpha}w_{\mathrm{th}}}{2}\left[\frac{1 + \frac{\gamma_D}{2}|\tilde{\mathcal{E}}_0|^2}{1 + i\delta\omega T_2}\mathcal{E}_+\right.$$
$$+ \chi^{(3)}|\tilde{\mathcal{E}}_0|^2\mathcal{E}_+\int\frac{dz}{L}|\Upsilon_0(z)|^2|\Upsilon_+(z)|^2$$
$$\left. + \chi^{(3)}\tilde{\mathcal{E}}_0^2\mathcal{E}_-^*\int\frac{dz}{L}\Upsilon_0(z)^2\Upsilon_-^*(z)\Upsilon_+^*(z)\right], \quad (G30)$$

and a similar equation can be written for $\dot{\mathcal{E}}_-$. We define the longitudinal overlap integrals

$$\Gamma_{\mathrm{self}} = \int_0^L \frac{dz}{L}|\Upsilon_0(z)|^2|\Upsilon_+(z)|^2 = 1, \quad (G31)$$

$$\Gamma_{\mathrm{cross}} = \int_0^L \frac{dz}{L}\Upsilon_0(z)^2\Upsilon_-^*(z)\Upsilon_+^*(z) = 1/2. \quad (G32)$$

The implication is that the self-mixing interaction of a sideband with itself, mediated by the primary mode intensity, is twice as large as the cross-mixing interaction of one sideband generating gain for the other sideband, again mediated by the primary mode intensity. This is true only for the cosine-shaped modes that we have assumed, and the overlap integrals will change when the longitudinal spatial profile changes, as when

the non-unity reflectivity of the facets is taken into account. The sideband equations become

$$\dot{\mathcal{E}}_+ = -\frac{\bar{\ell}}{2}\mathcal{E}_+ + \frac{\bar{\alpha}\,w_{\mathrm{th}}}{2}\left[\frac{1+\frac{\gamma_D}{2}|\tilde{\mathcal{E}}_0|^2}{1+i\delta\omega T_2}\mathcal{E}_+\right.$$
$$\left. + \Gamma_{\mathrm{self}}\chi^{(3)}|\tilde{\mathcal{E}}_0|^2\mathcal{E}_+ + \Gamma_{\mathrm{cross}}\chi^{(3)}\tilde{\mathcal{E}}_0^2\mathcal{E}_-^*\right], \quad (G33)$$

$$\dot{\mathcal{E}}_- = -\frac{\bar{\ell}}{2}\mathcal{E}_- + \frac{\bar{\alpha}\,w_{\mathrm{th}}}{2}\left[\frac{1+\frac{\gamma_D}{2}|\tilde{\mathcal{E}}_0|^2}{1-i\delta\omega T_2}\mathcal{E}_-\right.$$
$$\left. + \Gamma_{\mathrm{self}}\chi^{(3)*}|\tilde{\mathcal{E}}_0|^2\mathcal{E}_- + \Gamma_{\mathrm{cross}}\chi^{(3)*}\tilde{\mathcal{E}}_0^2\mathcal{E}_+^*\right], \quad (G34)$$

which we express as

$$\begin{pmatrix}\dot{\mathcal{E}}_+\\ \dot{\mathcal{E}}_-^*\end{pmatrix} = \begin{pmatrix}M_+ & R_+\\ R_-^* & M_-^*\end{pmatrix}\begin{pmatrix}\mathcal{E}_+\\ \mathcal{E}_-^*\end{pmatrix}, \quad (G35)$$

where

$$M_+ = M_-^* = -\frac{\bar{\ell}}{2} + \frac{\bar{\alpha}\,w_{\mathrm{th}}}{2}\left(\frac{1+\frac{\gamma_D}{2}|\tilde{\mathcal{E}}_0|^2}{1+i\delta\omega T_2} + \Gamma_{\mathrm{self}}\chi^{(3)}|\tilde{\mathcal{E}}_0|^2\right), \quad (G36)$$

$$R_+ = R_-^* = \frac{\bar{\alpha}\,w_{\mathrm{th}}}{2}(\Gamma_{\mathrm{cross}}\chi^{(3)}|\tilde{\mathcal{E}}_0|^2). \quad (G37)$$

As we did for the traveling-wave laser, the sideband gain is easily calculated from these two coupled first-order differential equations. Normalizing the gain to the total loss, we find

$$\frac{\bar{g}}{\bar{\ell}} = \frac{1+\frac{\gamma_D}{2}|\tilde{\mathcal{E}}_0|^2}{1+(\delta\omega T_2)^2} + \mathrm{Re}[\chi^{(3)}]|\tilde{\mathcal{E}}_0|^2$$
$$\times \begin{cases}\Gamma_{\mathrm{self}} + \Gamma_{\mathrm{cross}} = \frac{3}{2} & (\mathrm{AM})\\ \Gamma_{\mathrm{self}} - \Gamma_{\mathrm{cross}} = \frac{1}{2} & (\mathrm{FM})\end{cases}, \quad (G38)$$

which is equivalent to Eq. (19) after expressing $|\tilde{\mathcal{E}}_0|^2$ in terms of $p$ from Eq. (E13). There are two things to notice here. As the laser pumping is increased, the term $\gamma_D|\tilde{\mathcal{E}}_0|^2/2$ grows, and consequently the gain is not clamped at the threshold value. This is due to spatial hole burning or, more precisely, the imperfect overlap of the standing-wave modes together with a finite amount of carrier diffusion. We view this background gain as a Lorentzian shape whose amplitude increases with the pumping, and is therefore fully capable of pulling the sidebands above threshold, without any additional PP contribution to the gain.

Second, the PP contribution to the gain never vanishes. Even when the sidebands are phased such that an FM waveform is emitted from the laser, there is still a PP within the laser cavity. The reason for this is the imperfect overlap of the two sidebands' spatial modes, which means that at any given position within the cavity, the plus and minus sidebands are likely to have different amplitudes. Therefore, even if the two sidebands are phased such that their contributions to the beat note at $\delta\omega$ destructively interfere with each other, the destruction is not perfect. The amplitude of the PP varies with position in the cavity, and in locations where the two sideband amplitudes are equal the PP will not exist, but the spatially averaged effect of the FM PP yields the factor of $\frac{1}{2}$ in Eq. (G38). By the same token, sidebands phased for AM will
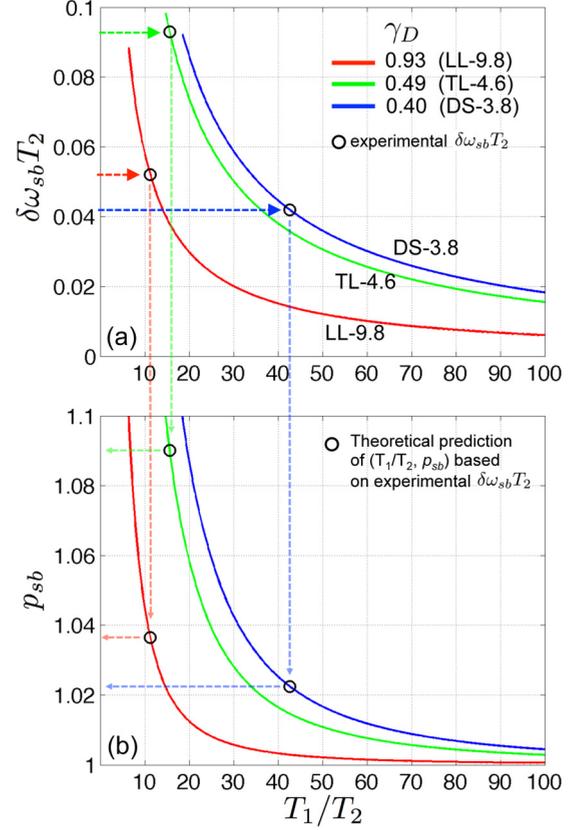


FIG. 12. Numerical solutions of the instability threshold obtained by setting the gain $\bar{g}$ in Eq. (G38) equal to the loss $\bar{\ell}$, yielding both (a) the sideband separation $\delta\omega_{\mathrm{sb}}T_2$ and (b) the pumping $p_{\mathrm{sb}}$. The experimentally measured values of $\delta\omega_{\mathrm{sb}}T_2$ are compared to the theory to infer $T_1/T_2$, which also gives the theoretical prediction for the instability threshold $p_{\mathrm{sb}}$.

not fully constructively interfere, yielding a factor of $\frac{3}{2}$ for the PP contribution to the gain rather than the factor 2, as it would be for the traveling-wave laser.

Depending on the particular values of $T_1$, $T_2$, and $\gamma_D$, either FM or AM sidebands will have a lower instability threshold. This can be determined by solving Eq. (G38) for the pumping level at which the $\bar{g} = \bar{\ell}$, and gives rise to the three different kinds of instability discussed in Sec. III D of the main text.

## APPENDIX H: COMPARISON OF THEORY AND DATA

In this Appendix, we take the theory at face value and calculate the predictions of the theory for the three uncoated lasers. We then compare the results with the measurements. (In the main text, we focused on comparing the experimental $\delta\omega_{\mathrm{sb}}$ with the theoretical $\delta\omega_{\mathrm{cr}}$, a comparison that is more robust against our uncertainty in $\gamma_D$ and $T_1$ and neglect of GVD and $\Delta$.)

Because the theory assumes end mirrors with unity reflectivity, we can only expect Eq. (G38) to apply reasonably well to the uncoated QCLs. For each device, $\gamma_D$ is calculated using the theoretical value of $T_{\mathrm{up}}$ (calculated from the band structure) and the diffusion constant $D = 77$ cm$^2$/s [49], giving $\gamma_D = 0.4$ (DS-3.8), 0.49 (TL-4.6), and 0.93 (LL-9.8). For these large values of $\gamma_D$, the incoherent gain increases rapidly with the

TABLE II. Summary of the input parameters and the output theoretical predictions for the three uncoated devices. $T_{up}$ is used to calculate $\gamma_D$. $T_2$ and $\delta\omega_{sb}T_2$ are measured quantities, from which the columns $T_1/T_2$ and $p_{sb}$ are calculated from the theory, as explained in Fig. 12. The quantity $p_{sb}$ clearly underestimates the measured $J_{sb}/J_{th}$ for reasons discussed in the text.

| Device | $T_{up}$ (ps) | $\gamma_D$ | $T_2$ (fs) | $\delta\omega_{sb}T_2$ | $T_1/T_2$ | $p_{sb}$ | $J_{sb}/J_{th}$ |
|--------|-----------|---------|---------|------------------|---------|--------|-------------|
| LL-9.8 | 0.54 | 0.93 | 81 | 0.052 | 11 | 1.04 | 1.14 |
| TL-4.6 | 1.7 | 0.49 | 74 | 0.093 | 16 | 1.09 | 1.17 |
| DS-3.8 | 1.74 | 0.40 | 43 | 0.042 | 43 | 1.02 | 1.12 |

pumping, and we find from Eq. (G38) that the FM instability will have a lower threshold than the AM instability, regardless of the value of $T_1$. The gain recovery time $T_1$ of each QCL is not as easily calculable as $T_{up}$ because it depends on a few other time constants of the active region, such as the escape time of the electron from one injector region to the next active region. Therefore, we treat $T_1$ as a variable and calculate the instability threshold $p_{sb}$ and sideband spacing $\delta\omega_{sb}$ as a function of $T_1$. The resulting curves are shown in Fig. 12, and a summary of all input and output parameters is given in Table II. By comparing the curves with the measured values

of $\delta\omega_{sb}$, we can deduce the values $T_1 = 1.83$ ps (DS-3.8), 1.15 ps (TL-4.6), and 0.91 ps (LL-9.8). For these values of $T_1$, the theory predicts an instability threshold of $p_{sb} = 1.02$ (DS-3.8), 1.09 (TL-4.6), and 1.04 (LL-9.8). It is encouraging that these fitted values of $T_1$ are close to the accepted value of the QCL gain recovery time, which has been shown by pump-probe experiments [62,63] and theory [64] to be around 2 ps. However, the predicted $p_{sb}$ is significantly lower than the measured values $J_{sb}/J_{th} = 1.12$ (DS-3.8), 1.17 (TL-4.6), and 1.14 (LL-9.8), and the discrepancy is made worse by the fact that $J/J_{th}$ is likely an underestimate of $p$ (see the discussion in Appendix A). The fact that the theory underestimates the instability threshold is perhaps not surprising, as we have only made sure that one of the two necessary conditions for sideband oscillation is satisfied (gain, not phase). Our neglect of the current inhomogeneity also contributes to the underestimation, as discussed in Appendix E. We hope that future work which accounts for the detuning $\Delta$, the detuning between the lasing mode and the cold-cavity mode it occupies, the GVD, and the current inhomogeneity can accurately predict the instability threshold, which would be a milestone in the understanding of lasers, and also yield a novel laser characterization method of lifetimes and diffusion rates by comparing measured values of $p_{sb}$ and $\delta\omega_{sb}$ to an established theory.

[1] T. J. Kippenberg, S. M. Spillane, and K. J. Vahala, Phys. Rev. Lett. **93**, 083904 (2004).

[2] A. B. Matsko, A. A. Savchenkov, D. Strekalov, V. S. Ilchenko, and L. Maleki, Phys. Rev. A **71**, 033804 (2005).

[3] P. Del'Haye, T. Herr, E. Gavartin, M. L. Gorodetsky, R. Holzwarth, and T. J. Kippenberg, Phys. Rev. Lett. **107**, 063901 (2011).

[4] T. Herr, K. Hartinger, J. Riemensberger, C. Y. Wang, E. Gavartin, R. Holzwarth, M. L. Gorodetsky, and T. J. Kippenberg, Nat. Photonics **6**, 480 (2012).

[5] T. Herr, V. Brasch, J. D. Jost, C. Y. Wang, N. M. Kondratiev, M. L. Gorodetsky, and T. J. Kippenberg, Nat. Photonics **8**, 145 (2013).

[6] K. Saha, Y. Okawachi, B. Shim, J. S. Levy, R. Salem, A. R. Johnson, M. A. Foster, M. R. Lamont, M. Lipson, and A. L. Gaeta, Opt. Express **21**, 1335 (2013).

[7] W. E. Lamb, Phys. Rev. **134**, A1429 (1964).

[8] H. Haken, Z. Phys. **190**, 327 (1966).

[9] H. Risken, C. Schmid, and W. Weidlich, Z. Phys. **194**, 337 (1966).

[10] H. Risken and K. Nummedal, J. Appl. Phys. **39**, 4662 (1968).

[11] R. Graham and H. Haken, Z. Phys. **213**, 420 (1968).

[12] S. Hendow and M. Sargent, Opt. Commun. **40**, 385 (1982).

[13] S. T. Hendow and M. I. Sargent, Opt. Commun. **43**, 59 (1982).

[14] L. W. Hillman, R. W. Boyd, and C. R. Stroud, Opt. Lett. **7**, 426 (1982).

[15] L. A. Lugiato, L. M. Narducci, D. K. Bandy, and N. B. Abraham, Opt. Commun. **46**, 115 (1983).

[16] L. W. Hillman, J. Krasinski, R. W. Boyd, and C. R. Stroud, Phys. Rev. Lett. **52**, 1605 (1984).

[17] L. A. Lugiato, L. M. Narducci, E. V. Eschenazi, D. K. Bandy, and N. B. Abraham, Phys. Rev. A **32**, 1563 (1985).

[18] L. W. Hillman, J. Krasinski, K. Koch, and C. R. Stroud, J. Opt. Soc. Am. B **2**, 211 (1985).

[19] R. S. Gioggia and N. B. Abraham, Phys. Rev. Lett. **51**, 650 (1983).

[20] S. Autler and C. Townes, Phys. Rev. **100**, 703 (1955).

[21] B. Senitzky, G. Gould, and S. Cutler, Phys. Rev. **130**, 1460 (1963).

[22] B. Senitzky, IEEE Trans. Microw. Theory Tech. **MTT-16**, 728 (1968).

[23] M. C. Newstein, Phys. Rev. **167**, 89 (1968).

[24] B. Mollow, Phys. Rev. **188**, 1969 (1969).

[25] C. R. Stroud, Phys. Rev. A **3**, 1044 (1971).

[26] F. Schuda, C. R. Stroud, and M. Hercher, J. Phys. B: At. Mol. Phys. **7**, L198 (1974).

[27] F. Y. Wu, R. E. Grove, and S. Ezekiel, Phys. Rev. Lett. **35**, 1426 (1975).

[28] H. Carmichael and D. Walls, J. Phys. B: At. Mol. Phys. **9**, 1199 (1976).

[29] C. Cohen-Tannoudji and S. Reynaud, J. Phys. B: At. Mol. Phys. **10**, 345 (1977).

[30] R. W. Boyd, M. G. Raymer, P. Narum, and D. J. Harter, Phys. Rev. A **24**, 411 (1981).

[31] D. J. Harter and R. W. Boyd, IEEE J. Quantum Electron. **16**, 1126 (1980).

[32] L. A. Lugiato, F. Prati, D. K. Bandy, L. M. Narducci, P. Ru, and J. Tredicce, Opt. Commun. **64**, 167 (1987).

[33] A. Gordon, C. Y. Wang, L. Diehl, F. X. Kärtner, A. Belyanin, D. Bour, S. Corzine, G. Höfler, H. C. Liu, H. Schneider, T. Maier, M. Troccoli, J. Faist, and F. Capasso, Phys. Rev. A **77**, 053804 (2008).

[34] R. L. Tober, J. D. Bruno, S. Suchalkin, and G. Belenky, J. Opt. Soc. Am. B **31**, 2399 (2014).

[35] Q. Y. Lu, M. Razeghi, S. Slivken, N. Bandyopadhyay, Y. Bai, W. J. Zhou, M. Chen, D. Heydari, A. Haddadi, R. McClintock, M. Amanti, and C. Sirtori, Appl. Phys. Lett. **106**, 051105 (2015).

[36] M. C. Soriano, J. García-Ojalvo, C. R. Mirasso, and I. Fischer, Rev. Mod. Phys. **85**, 421 (2013).

[37] A. Hugi, G. Villares, S. Blaser, H. C. Liu, and J. Faist, Nature (London) **492**, 229 (2012).

[38] G. Villares, A. Hugi, S. Blaser, and J. Faist, Nat. Commun. **5**, 5192 (2014).

[39] G. Villares, S. Riedi, J. Wolf, D. Kazakov, M. J. Süess, M. Beck, and J. Faist, Optica **3**, 252 (2016).

[40] J. B. Khurgin, Y. Dikmelik, A. Hugi, and J. Faist, Appl. Phys. Lett. **104**, 081118 (2014).

[41] G. Villares and J. Faist, Opt. Express **23**, 1651 (2015).

[42] H. Li, P. Laffaille, D. Gacemi, M. Apfel, C. Sirtori, J. Leonardon, G. Santarelli, M. Rösch, G. Scalari, M. Beck, J. Faist, W. Hänsel, R. Holzwarth, and S. Barbieri, Opt. Express **23**, 33270 (2015).

[43] V. E. Zakharov and L. A. Ostrovsky, Phys. D (Amsterdam) **238**, 540 (2009).

[44] A. Schliesser, N. Picque, and T. W. Hansch, Nat. Photonics **6**, 440 (2012).

[45] D. Hofstetter, M. Beck, T. Aellen, J. Faist, U. Oesterle, M. Ilegems, E. Gini, and H. Melchior, Appl. Phys. Lett. **78**, 1964 (2001).

[46] J. Faist, D. Hofstetter, M. Beck, T. Aellen, M. Rochat, and S. Blaser, IEEE J. Quantum Electron. **38**, 533 (2002).

[47] F. Xie, C. Caneau, H. P. Leblanc, N. J. Visovsky, S. C. Chaparala, O. D. Deichmann, L. C. Hughes, C. Zah, D. P. Caffey, and T. Day, IEEE J. Sel. Top. Quantum Electron. **17**, 1445 (2011).

[48] D. Hofstetter and J. Faist, IEEE Photonics Technol. Lett. **11**, 1372 (1999).

[49] J. Faist, *Quantum Cascade Lasers*, 1st ed. (Oxford University Press, Oxford, 2013).

[50] J. Faist, G. Villares, G. Scalari, M. Rösch, C. Bonzon, A. Hugi, and M. Beck, Nanophotonics **5**, 272 (2016).

[51] L. Allen and J. Eberly, *Optical Resonance and Two-Level Atoms* (Dover, New York, 1987).

[52] A. Siegman, *Lasers* (University Science Books, Sausalito, 1986), pp. 473–489.

[53] A. K. Wójcik, P. Malara, R. Blanchard, T. S. Mansuripur, F. Capasso, and A. Belyanin, Appl. Phys. Lett. **103**, 231102 (2013).

[54] Y. Wang and A. Belyanin, Opt. Express **23**, 4173 (2015).

[55] Y. K. Chembo, D. V. Strekalov, and N. Yu, Phys. Rev. Lett. **104**, 103902 (2010).

[56] N. Vukovic, J. Radovanovic, V. Milanovic, and D. L. Boiko, arXiv:1601.03212.

[57] H. G. Danielmeyer, J. Appl. Phys. **42**, 3125 (1971).

[58] R. W. Boyd, *Nonlinear Optics*, 2nd ed. (Academic, San Diego, 2003).

[59] D. A. Kleinman, Phys. Rev. **174**, 1027 (1968).

[60] N. J. Holonyak, R. M. Kolbas, R. D. Dupuis, and P. D. Dapkus, IEEE J. Quantum Electron. **QE-16**, 170 (1980).

[61] T. S. Mansuripur, G. M. De Naurois, A. Belyanin, and F. Capasso, Optica **2**, 48 (2015).

[62] H. Choi, L. Diehl, Z. K. Wu, M. Giovannini, J. Faist, F. Capasso, and T. B. Norris, Phys. Rev. Lett. **100**, 167401 (2008).

[63] H. Choi, L. Diehl, Z. K. Wu, M. Giovannini, J. Faist, F. Capasso, and T. B. Norris, IEEE J. Quantum Electron. **45**, 307 (2009).

[64] M. A. Talukder, J. Appl. Phys. **109**, 033104 (2011).

[65] Y. K. Chembo and N. Yu, Phys. Rev. A **82**, 033801 (2010).