Title
Design strategy for terahertz quantum dot cascade lasers

Permalink
https://escholarship.org/uc/item/6jm701w1

Journal
OPTICS EXPRESS, 24(22)

ISSN
1094-4087

Authors
Burnett, BA
Williams, BS

Publication Date
2016-10-31

DOI
10.1364/OE.24.025471

Peer reviewed
Design strategy for terahertz quantum dot cascade lasers

BENJAMIN A. BURNETT AND BENJAMIN S. WILLIAMS

Department of Electrical Engineering and California NanoSystems Institute, University of California, Los Angeles, Los Angeles, California 90095, USA
*bburnett@ucla.edu

Abstract: The development of quantum dot cascade lasers has been proposed as a path to obtain terahertz semiconductor lasers that operate at room temperature. The expected benefit is due to the suppression of nonradiative electron-phonon scattering and reduced dephasing that accompanies discretization of the electronic energy spectrum. We present numerical modeling which predicts that simple scaling of conventional quantum well based designs to the quantum dot regime will likely fail due to electrical instability associated with high-field domain formation. A design strategy adapted for terahertz quantum dot cascade lasers is presented which avoids these problems. Counterintuitively, this involves the resonant depopulation of the laser’s upper state with the LO-phonon energy. The strategy is tested theoretically using a density matrix model of transport and gain, which predicts sufficient gain for lasing at stable operating points. Finally, the effect of quantum dot size inhomogeneity on the optical lineshape is explored, suggesting that the design concept is robust to a moderate amount of statistical variation.

© 2016 Optical Society of America

OCIS codes: (140.5965) Semiconductor lasers, quantum cascade; (250.5590) Quantum-well, -wire and -dot devices; (140.3070) Infrared and far-infrared lasers.

References and links
1. Introduction

Terahertz quantum cascade lasers (QCLs) are increasingly viable coherent sources in the 1-5 THz range, having now advanced to pulsed peak power above 1 W [1] and CW power above 100 mW [2, 3]. However, room-temperature operation remains elusive with the maximum operating temperature at only 200 K [4]. The primary reason for this limitation is now generally accepted to be nonradiative relaxation of electrons in the upper radiative state due to thermally-activated electron-longitudinal-optical (LO-) phonon scattering [5-7]. In a THz QC-laser, the energy spacing between the upper radiative state “U” and lower radiative state “L” is smaller than the LO-phonon energy: $E_{UL} < E_{LO}$ (36 meV in GaAs = 8.7 THz). As the electronic temperature increases, electrons gain sufficient in-plane kinetic energy to emit LO-phonons; this
process rapidly depletes the upper state and prevents lasing at room temperature.

It has been proposed that replacing the quantum wells in a QCL with quantum dots could offer a solution due to the discretization of the electronic subbands into “sublevels.” In principle this would lead to a so-called “phonon bottleneck,” where LO-phonon scattering is suppressed between pairs of sublevels provided they are not resonant with $E_{LO}$ [8–10]. The existence of this effect has been experimentally confirmed by directly measuring THz intersublevel relaxation times as long as 1 ns in self-assembled quantum dots at 10 K (extrapolated to tens of picoseconds at room temperature) [11], as well as the raising of a THz QCL’s operating temperature from 160 K to 225 K by applying a strong magnetic field to discretize the subbands into Landau levels [12]. The candidate geometries for such a quantum dot cascade laser (QDCL) include self-assembled quantum dots [13, 14], nanopillars etched from the top-down into QCL active material [15–17], and nanopillars epitaxially grown from the bottom-up [18, 19].

Due to the challenges of developing the necessary fabrication and growth techniques, simulation is particularly important to validate the QDCL concept and to guide design. A proper theoretical treatment cannot rely on the usual Fermi’s Golden Rule rate equation approaches, because the electron-LO-phonon interaction takes on an entirely new character owing to the formation of strongly-coupled intersublevel electron-phonon polarons [24]. So far, both nonequilibrium Green’s function (NEGF) and density matrix models have predicted significant population inversion and THz gain at room temperature, especially once the high confinement regime is reached (lateral dimensions of 20-30 nm depending on material) so that only the lowest lateral state plays a role in transport [20–22]. However, in Refs. [21, 22], an extreme subthreshold parasitic current channel was predicted when the injector sublevel $I$ from the previous module aligns with the lower radiative state $L$. This effect would create a negative differential resistance (NDR) bias regime beyond the parasitic point. The NDR regime leads to space charge buildup and high-field domain formation, which effectively renders the intended operating bias point inaccessible, and prevents laser operation.

NDR occurs because transport is far more efficient at the parasitic bias compared to the higher design bias. This is a natural consequence of the lower radiative state having a short lifetime due to relaxation by LO-phonon emission: this occurs by design by setting the lower subband energy such that $E_{LI} \approx E_{LO}$. In conventional QCLs, this parasitic channel is kept modest by using thick injection barriers to ensure that the tunneling process — and not the lifetime $\tau_L$ — is the bottleneck to current flow [23]. In the QDCL, this strategy will not work, because the longer lifetimes and the longer dephasing times mean that barriers would have to be impractically large to suppress the parasitic channel.

In this paper, we present a counterintuitive strategy for the design of a THz QDCL which avoids the expected problem of electrical instability. This is in contrast to previous theoretical investigations which have for the most part focused upon designs which follow the same design rules as conventional QCLs. The central concept is to engineer the sublevel spacings so that resonant-phonon emission occurs from the laser’s upper state, i.e. $E_{UI} \approx E_{LO}$, rather than from the lower state as is widely adopted in ordinary QCLs. This has the effect of ensuring the maximum current occurs at the design bias, rather than at below. The strategy is tested theoretically using the density matrix model presented in [22], extended to a level of generality so that both the original and new designs are tested using the same code. It is predicted that sufficient gain is available at room temperature and at electrically stable operating points, although the gain level and saturation intensity are lower than for the conventional strategy. Finally, simulations of quantum dot size inhomogeneity are performed, which suggest that some amount of size fluctuation does not necessarily prohibit lasing.

The paper is organized as follows: Section 2 gives an overview of the design concept including the physics of electron-LO-phonon interaction in quantum dots, Section 3 overviews the density matrix model used to assess the new strategy, Section 4 gives the calculated results, Section 5
addresses the effect of quantum dot size inhomogeneity, and concluding remarks are made in Section 6.

2. Design concept

2.1. Polaron physics

A brief overview of the relevant polaron physics is given here; for a more detailed discussion the reader is referred to [22]. The discrete nature of electronic sublevels in quantum dots combined with the nearly dispersionless LO-phonon spectrum results in a strong coupling of the two and the formation of electron-phonon quasiparticles known as “polarons”. Because the combined system does not have a continuum of states, irreversible decay does not occur. Instead, the electron continually emits and re-absorbs an LO-phonon, resulting in a Rabi oscillation which continues until the decay of the LO-phonon or another interruption [24, 25].

It can be shown that a given intersublevel electronic transition emits phonons into only a single mode spatially localized around the transition. The polaron coupling strength to this mode is defined as $\Omega_{pol}$, which couples for example the product states $|\psi_2; 0\rangle$ and $|\psi_1; 1\rangle$, which are formed from the electron states $\psi_n$ and phonon mode occupations $m$. This is illustrated qualitatively in Fig. 1(a), where we consider the energy eigenstates obtained by diagonalization of the electron-phonon Hamiltonian in a fictitious InAs quantum dot of variable dimension such that the intersublevel energy spacing $E_2 - E_1$ is varied. The dashed lines are the state energies without the electron-phonon interaction; inclusion of the interaction causes an anticrossing behavior between the polaron eigenstates. Full hybridization occurs for an electronic transition tuned to $E_{LO}$ (29 meV in InAs), whereas for detuned transitions the eigenstates return more so to their separate “bare” characters, i.e. “electron-like” or “phonon-like.”

LO-phonons can irreversibly decay into various pairs of acoustic phonons, as long as the total momentum and energy of the pair is conserved. The overall relaxation rate can be computed by integrating over the continuum of possible pairs, and has been performed in various works [25,26]; here we borrow the results from [26] and approximate a 2.5 ps relaxation time to equilibrium at room temperature. From this rate, the transition rates between the phonon number (Fock) states of a given mode, which are the decay and generation of phonons, are deduced.

Fig. 1(b) shows the approximate relaxation dynamics for three different intersublevel energies: $E_{21} = 10, 20,$ and $29$ meV. Two key characteristics are observed. First, despite the existence of
Fig. 2. Electronic bandstructures and level schematics at parasitic/design biases for (a) a conventional QCL design and (b) a modified design for a QDCL. The chosen material system is InAs/InAsP and the layer thickness in nm starting with the tunnel barrier are 5.5/14/3/30 for (a) and 5.5/18/3/35 for (b), so that the only difference is a widening of the wells. The level schematics illustrate the design strategy which is to engineer resonant LO-phonon depopulation of the upper state $\Psi_U$, rather than the lower state $\Psi_L$, to the injector state $\Psi_I$.

Rabi oscillations, the envelope of the relaxation process can be described by an approximate lifetime. Second, this lifetime increases for smaller $E_{21}$, as the intersublevel transition becomes further detuned from $E_{LO}$, and the polaron becomes less "phonon-like". Fig. 1(c) repeats the example of Fig. 1(b), only now including a finite pure dephasing time $T^*_2$. $T^*_2$ in quantum dots has been investigated experimentally using a four-wave mixing experiment in [27], where it was found to decrease from approximately 90 ps at 10 K to 9 ps at 120 K. Calculations in [28] indicate that $T^*_2$ is intricately dependent on the details of the energy structure, but expected values are a few ps at 300 K. In this work we will use $T^*_2 = 3$ ps, phenomenologically, which is an order of magnitude longer than typical values assumed for $T^*_2$ in quantum wells. As seen in Fig. 1, the presence of pure dephasing has a fairly small effect on transport, although the effect on the optical linewidth is more significant; further discussion is found in [22].

2.2. Design strategy

The most successful THz QCL designs are based on the “resonant-phonon” concept, which is shown in Fig. 2(a) for a simple 3-level laser design (adapted from [22, 33]), with upper radiative, lower radiative, and injector states labeled $U$, $L$, and $I$ respectively. Fast depopulation is ensured by designing $E_L - E_I \approx E_{LO}$.

We propose the new concept specifically for QDCL design where it is the upper, rather than the lower, state that is placed to depopulate at resonance with LO-phonons. In other words, the design is such that $E_U - E_I \approx E_{LO}$. This is shown in the design of Fig. 2(b), which is adapted from the previous design simply by widening the wells, in contrast to the suggestion in [21] of widening the barriers. This would certainly not be a useful concept in ordinary QCL design, as the upper state lifetime would most likely be far too short to support sufficient population inversion at a reasonable current density. However, here it is imperative that the current maximum occurs at the intended design bias (where $I'$ injects into $U$) rather than at a lower bias (where
injects into $L$). And indeed, since $E_L - E_I < E_{LO}$, depopulation is slowed. However, all of these things notwithstanding, it will be shown that the laser can in fact still support a sufficient population even under the condition that the effective lifetime of the upper state is shorter than the lower.

The condition for electrical stability is that the current at design bias $J_0$ should be greater than the current at the parasitic resonance $J_{para}$. Although a real estimation of the current at either point requires a detailed simulation, the design strategy that we employ can be understood by thinking of the polaron relaxations as leading to effective relaxation lifetimes $\tau_U$ and $\tau_L$ (similar to shown in Fig. 1). Since the primary bottleneck to transport in a QDCL is the polaron relaxation, the injector level can be expected to share population with the upper/lower state when biased to resonance with either, if the further approximation of perfectly selective injection is made. This leads to an estimation of the currents as:

$$J_{para} \approx \frac{N_s q}{2\tau_L}, \quad J_0 \approx \frac{N_s q}{2\tau_U}, \quad \text{(1)}$$

where $N_s$ is the sheet doping density. It is understood through Eq. 1 that the condition $\tau_U < \tau_L$ is required for electrical stability.

Furthermore, the concept of rates would give the steady-state population inversion in our system as:

$$\Delta N = \frac{J}{q} \left(1 - \frac{\tau_L}{\tau_{UL}}\right), \quad \text{(2)}$$

which establishes the simple condition that the lower state lifetime $\tau_L$ need only be shorter than the transition time across the radiative states $\tau_{UL}$ to yield positive population inversion. A short upper-state lifetime $\tau_U$ reduces the magnitude of the population inversion, but importantly does not change the sign. We expect that the condition $\tau_{UL} > \tau_L$ can be satisfied in our system as long as the depopulation energy $E_L - E_I$ is tuned closer to $E_{LO}$ than the radiative energy $E_U - E_L$.

3. Density matrix transport model

Transport models in QCLs are often based on a density matrix, because of the fact that scattering can localize electrons into states whose description requires subband correlations. The general approach is to describe certain evolution as coherent, driven by a Hamiltonian $H$ through the Liouville-von Neumann equation $i\hbar \dot{\rho} = -[H, \rho]$ ($\rho$ is the density matrix), while other parts of the evolution are described as incoherent scattering. Effects of tunneling can be recovered by making a phenomenological basis choice localized on either side of the barrier, and incorporating transitions and dephasing within this basis [29], or by computing the scattering more generally where all density matrix elements can be coupled [30, 31]. In the simpler density matrix models, which can be traced all the way back to the first proposal for a QCL [32], analytic solutions are derived to give useful equations describing parameter dependencies [33, 34], whereas more recently numerical solutions have been used to simulate more complex systems. This includes the generalized scattering approaches which give insight into optical pulse response [30] and in-plane behavior [31], and also phenomenological approaches in which the steady-state under coherent, nonperturbative optical excitation is solved for to yield gain spectra [35] and optical nonlinearities [36].

In the density matrix model used here, described in more detail in [22], the basis of states is formed by the tensor product of the electronic states in each repetitive module and the occupations of the relevant phonon modes. Thus, the coherent part of the evolution is composed of the state energies, tunneling, electron-phonon interaction, and optical field; the incoherent evolution includes the phonon decay and generation, and the pure dephasing. Periodic boundary conditions are applied. The Hamiltonian and density matrix are allowed frequency content at dc and the
As was the case in [22], we are again assuming that the pillar diameter is sufficiently small that only the lateral ground states are relevant to the transport problem. The pillar cross-section is approximated as circular with a diameter of 30 nm, which results in a separation of 66 meV between the ground and first excited lateral states in InAs, well above the LO-phonon energy of 29 meV. The phonon interaction is described using three orthogonalized phonon modes which span the space defined by the three possible intramodule transitions: $U-L$, $U-I$, and $L-I$. This is an extension of the simulation in Ref. [22], which only included two modes, and allows for both designs to be simulated using the exact same code. The maximum total occupation in all modes is allowed to reach up to two — sufficient to capture the relevant phenomena at room temperature. A schematic of the full electron-phonon product space similar to Fig. 4 in [22] can be drawn, but with the addition of the extra phonon mode, it is too complicated to reproduce here.

The modeling in this work is worth comparing to the NEGF approaches [20, 21]. The NEGF calculations have the advantage over density matrix models of not requiring a phenomenological basis choice for the electron wavefunctions; however, the complexity of NEGF presents somewhat of a restriction that only the electron can practically be tracked. This means that the presence of the LO-phonons enters into the calculation as an averaged field which acts as a bath on the electrons and is assumed to remain at equilibrium, thus possibly missing effects where generated nonequilibrium phonons act back on the electron system. A density matrix approach is computationally lighter than NEGF and thus allows for the LO-phonons to be treated on the same footing as the electrons, so that the role of the bath is pushed further down to the acoustic phonons [22].

4. Results

Simulated results of the transport characteristics and bias-dependent small-signal optical gain at 300 K are in Fig. 4. It is assumed for both designs that the pillars are spaced on a square lattice of 60 nm pitch, and that one electron exists per module, amounting to an overall electron density of approximately $5 \times 10^{15}$ cm$^{-3}$. Due to the very large CPU time required, these simulations make use of highly-parallel (high-throughput) computing.

The strong parasitic channel is evident in the conventional design at $\sim$25-30 mV/module, with a current density approximately $5 \times$ larger than at the design bias of 42 mV/module. In contrast, while the modified design does exhibit a parasitic current channel at 25 mV/module, the current is much larger at the design point of 29 mV/module. The conventional design is calculated to exhibit higher gain at design bias than the modified design, but the bias point is unstable making this gain entirely inaccessible. In fact, hardly any gain at all is predicted in the conventional design at any stable points. The modified design, on the other hand, still produces a significant amount of gain (on the order of 100 cm$^{-1}$) in the region of stability and around the frequency that is visible from the electronic bandstructure. This result suggests that our concept is valid.

Effects of gain saturation are investigated in Fig. 4. In these simulations, the small-signal gain at frequency $\omega$ is calculated under increased intensity at the frequency of peak small-signal gain $\omega_{\text{peak}}$, so that cross-saturation of $\omega$ by intensity at $\omega_{\text{peak}}$ is obtained. Extension of the model to include multiple frequencies was performed in manner similar to [36]. Results suggest that in both designs, the saturation is highly nontrivial owing to the complex nature of the polaron-coupled system; aspects of both homogeneous and inhomogenous saturation are visible. Importantly, spectral hole burning effects are important even for perfectly uniform quantum dot size. Fig. 4(c) shows the small-signal gain spectrum peak as a function of saturation by intensity at $\omega_{\text{peak}}$. The modified design exhibits a saturation intensity around the 10 W/mm$^2$ level, which
Fig. 3. Simulated transport characteristics and bias-dependent small-signal optical gain at 300 K for the conventional and modified designs, showing regions of electrical stability/instability. The designed and predicted gain in the modified design is circled by the dashed oval.

Fig. 4. Simulated effects of saturation on the conventional and modified designs at their respective design biases. Saturated gain profiles in (a) and (b) are cross-saturation by intensity at \( \omega_{\text{peak}} \) (marked by dashed line), and (c) is the peak of these profiles in relation to the small-signal gain at \( \omega_{\text{peak}} \).
is approximately one order of magnitude lower than the conventional design. As was done for the current (1) and gain (2), this can also be understood in terms of effective rates; the saturation intensity can be written for a three-level laser as:

\[ I_{\text{sat}}(\nu) = \frac{h\nu}{\sigma(\nu)\left[\tau_L + \tau_U\left(1 - \frac{\tau_L}{\tau_{UL}}\right)\right]} \]

(3)

where \(\nu\) is the optical frequency and \(\sigma(\nu)\) is the gain cross-section. The low saturation intensity is caused by the long lower state lifetime, which is a consequence of \(E_L - E_I < E_{LO}\).

5. Effects of inhomogeneity

Because quantum dot sizes cannot be controlled at the atomic precision level as is possible in quantum wells, size inhomogeneity can be expected to be an important factor in device operation, quite possibly dominating the optical linewidth. In the case of self-assembled quantum dots the growth process is inherently stochastic. Quantum dots defined by bottom up nanopillar growth and top-down etching have lateral dimensions defined by lithography — nonetheless some nonuniformity will likely persist. In all cases, due to the complexity of the gain profile which is split by the nontrivial interplay of tunneling and polaronic couplings, assessment of the inhomogeneous broadening cannot be simply assumed to follow a normal distribution.

To approximate the effect of inhomogeneity, simulations were run in which all length dimensions (4 layer widths and the diameter) were allowed to fluctuate by a Gaussian distribution of 5% standard deviation. The small-signal gain is computed for 50 variations on each design, which are all assumed to be biased at the design bias of the target geometry. This would ignore the effects of, for example, spatial correlation in dimensions that might be expected in the growth reactor, and the slightly different biases that each might see due to series combination. However, this still provides a rough look at the sensitivity of the gain profile to fluctuations in quantum dot dimensions. Results are shown in Fig. 5. The 5% standard deviation in lengths is expected to reduce the peak gain by approximately one half, although this would remain well above a typical cavity loss of 25 cm\(^{-1}\).

6. Conclusion

A useful strategy was presented for the design of a THz QDCL that avoids the predicted problem of electrical instability. The concept is, counterintuitively, to aim for resonant-phonon
depopulation of the laser’s upper, as opposed to lower, state. This creates a parallel current channel at design bias from $U \rightarrow I$ that ensures one can reach design bias. This can be considered conceptually equivalent to the strategy of using a shunt resistance to allow bias of a superlattice in the NDR regime so as to obtain Bloch gain [39]. Although somewhat lower gain is predicted in the modified design as compared to a conventional design hypothetically at design bias, this is still at the level of greater than 100 cm$^{-1}$ which should be easily sufficient for lasing. A downside brought about by our modification is an order-of-magnitude lower saturation intensity because of the long lower state lifetime, which could lead to a lower power level. Finally, simulations of quantum dot size inhomogeneity were performed, which suggests that at least a 5% uncertainty in the dot dimensions is tolerable. We should also note that this strategy will likely work only when $E_U - E_L < E_L - E_I$, such that the radiative transition is further from $E_{LO}$ than the depopulation. In InAs dots for example, this will limit the photon energy to at most 14 meV ($\sim$3.5 THz).

These results provide a concrete example of the contrast between successful QCL and QDCL design concepts, which is called for because of the inherently different transport physics. Further work could focus on more complicated designs involving a larger module, although this would be outside the practical capability of this modeling approach, or on the inclusion of charging effects [40].

Acknowledgements

This work was partially supported by NSF ECCS Grants 1150071 and 1509801. Computation was performed using the compute resources and assistance of the UW-Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy’s Office of Science.