



# Numerically efficient density-matrix technique for modeling electronic transport in mid-infrared quantum cascade lasers

S. Soleimanikahnoj<sup>1</sup> · O. Jonasson<sup>1</sup> · F. Karimi<sup>1</sup> · I. Knezevic<sup>1</sup>

Received: 27 August 2020 / Accepted: 21 November 2020 / Published online: 1 January 2021  
© Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

We present a numerically efficient density-matrix model applicable to mid-infrared quantum cascade lasers. The model is based on a Markovian master equation for the density matrix that includes in-plane dynamics, preserves positivity of the density matrix and does not rely on phenomenologically introduced dephasing times. Nonparabolicity in the band structure is accounted for with a three-band  $\mathbf{k} \cdot \mathbf{p}$  model, which includes the conduction, light-hole, and spin-orbit split-off bands. We compare the model to experimental results for QCLs based on lattice-matched as well as strain-balanced InGaAs/InAlAs heterostructures grown on InP. We find that our density-matrix model is in quantitative agreement with experiment up to threshold and is capable of reproducing results obtained using the more computationally expensive nonequilibrium Green's function formalism. We compare our density-matrix model to a semiclassical model where off-diagonal elements of the density matrix are ignored. We find that the semiclassical model overestimates the threshold current density by 29% for a 8.5-μm-QCL based on a lattice-matched heterostructure and 40% for a 4.6-μm-QCL based on a strain-balanced heterostructure, demonstrating the need to include off-diagonal density-matrix elements for accurate description of mid-infrared QCLs.

**Keywords** Quantum cascade laser · Density matrix · Markovian master equation · Dissipator · Quantum transport · Simulation · Modeling

## 1 Introduction

Quantum cascade lasers (QCLs) are unipolar light-emitting sources in the mid-infrared and terahertz parts of the electromagnetic spectrum [1, 2]. The gain medium in QCLs is a vertically-grown semiconductor heterostructure made of alternating layers of quantum wells and barriers. A population inversion between quasibound quantum states is achieved by a careful design of the layer structure and application of electrical bias. In this work, we focus on mid-infrared QCLs based on a InGaAs/InAlAs heterostructure grown on InP substrates. QCLs based on this material system have achieved room-temperature, continuous-wave, high-power operation in the mid-infrared range [3–5], and

have found applications in the fields of remote sensing [6] and metrology [7].

Numerical simulations play an essential role in the development of QCLs and can provide detailed insights into the microscopic details of QCL operation [8, 9]. Electron transport in quantum cascade lasers in both mid-infrared and terahertz regimes has been simulated via semiclassical (rate equations [10–13] and Monte Carlo [14–18]) and quantum techniques (density matrix [8, 19–30] and nonequilibrium Green's functions (NEGF) [31–38]). Semiclassical models are attractive due to their low computational costs, allowing inclusion of advanced phenomena such as nonequilibrium phonons [17]. However, semiclassical models involve the assumption that off-diagonal density-matrix (DM) elements (also referred to as the coherences) are much smaller than the diagonal ones (which correspond to level occupations). This approximation has been shown to fail in QCLs working in the terahertz [20, 39], as well as in the mid-infrared range [40].

NEGF simulations accurately capture quantum transport in QCLs [31, 33], but are computationally demanding. Density-matrix models offer a compromise between speed

✉ I. Knezevic  
iknezevic@wisc.edu

S. Soleimanikahnoj  
soleimanikah@wisc.edu

<sup>1</sup> Department of Electrical and Computer Engineering,  
University of Wisconsin-Madison, 1415 Engineering Dr.,  
Madison, WI 53706, USA

and accuracy. They are capable of describing partially coherent transport and require considerably less computational resources than NEGF. Previous theoretical work on QCLs using density-matrix approaches features various approximations: Simplified treatment of in-plane dynamics, where either a thermal distribution (Maxwell or Fermi-Dirac) is assumed [20, 34] or the density matrix is assumed to be separable into the cross-plain and in-plane directions (same in-plane distribution assumed for all subbands) [41], both of which may pose a problem because QCLs operate far from equilibrium, so in-plane distributions can deviate strongly from a thermal distribution and in-plane distributions can vary drastically between subbands [17, 34, 42]; semiclassical treatment within a subregion of a device (typically a single stage separated by thick injection barriers), while coupling between adjacent subregions is treated quantum mechanically using phenomenological dephasing times [20, 22, 34, 39]; a Redfield equation of motion that does not require the adoption of an artificially localized basis, which is important in THz QCLs [24]; a completely positive equation of motion for the density matrix with model dissipative terms of the Lindblad form and empirical parameters [30].

In this paper, we present a numerically efficient density-matrix model applicable to electronic transport in mid-infrared and terahertz QCLs. The model is based on a Markovian master equation for the density matrix that is of the so-called Lindblad form [29, 30, 43–45] and thus preserves the positivity of the density matrix. The model supplants the need for a phenomenological treatment of dissipation or dephasing: Electronic interaction with phonons, impurities, interface roughness, and random alloy disorder is described through the dissipator term derived directly from microscopic Hamiltonians. Full in-plane dynamics are included without assuming a thermal distribution or separation of the density matrix into in-plane and cross-plane components. Nonparabolicity in the band structure is treated using a three-band  $\mathbf{k} \cdot \mathbf{p}$  model. We note that the model is ideal for QCLs in which a finite number of reasonably well-localized states governs electronic transport, and is therefore better suited for mid-infrared than terahertz structures.

This paper is organized as follows. Section 2 contains an overview of the theory. We derive a Markovian master equation that governs the time evolutions of the single-electron density matrix and the linear response to a harmonic electric field. In Sect. 3, we discuss details regarding the numerical solution of the master equation and derive an iterative scheme to efficiently solve for the steady-state and linear-response density matrices. In Sect. 4, we verify the model by simulating mid-infrared QCLs based on InGaAs/InAlAs heterostructures grown on InP. We consider an 8.5  $\mu\text{m}$ -QCL based on a lattice-matched heterostructure and a 4.6- $\mu\text{m}$ -QCL based on a strain-balanced heterostructure. We compare our results to experiment, as well as to theoretical results based on NEGF

and to a semiclassical model where only diagonal values of the density matrix are considered. Concluding remarks are given in Sect. 5.

## 2 Theory

Electrons interacting with an environment can be described using a Hamiltonian on the form

$$H = H_0 + H_i + H_{\text{ph}}, \quad (1)$$

where  $H_0$  is the unperturbed electron Hamiltonian (in our case a semiconductor superlattice under an applied electric field),  $H_{\text{ph}}$  is the Hamiltonian of the environment (e.g., the free-phonon Hamiltonian), and  $H_i$  contains the interaction terms (e.g., electron–phonon interaction or scattering due to interface roughness). For transport calculations, we will use the eigenstates of  $H_0$  as a basis, which are calculated using a three-band  $\mathbf{k} \cdot \mathbf{p}$  Hamiltonian described in “Appendix 1”. We assume that the device area in the  $x$ – $y$  plane perpendicular to the transport ( $z$ ) direction is macroscopic with respect to the length of a single period and that the system behaves as translationally invariant in this plane. The eigenstates of  $H_0$  are labelled as  $|n, \mathbf{k}\rangle = |n\rangle \otimes |\mathbf{k}\rangle$ , where the discrete index  $n$  quantizes states in the transport direction ( $z$ ) and  $\mathbf{k}$  is the wave vector associated with free motion in the  $x$ – $y$  plane.

Owing to in-plane translational invariance, the reduced density matrix of electrons is diagonal in  $\mathbf{k}$  and only depends on the magnitude  $k = |\mathbf{k}|$ . The central quantities of interest are then the matrix elements  $\rho_{n,m}^{E_k}$ , which for  $n \neq m$  represent the coherence between eigenstates  $|n\rangle$  and  $|m\rangle$ , at the in-plane energy  $E_k$  while,  $\rho_{n,n}^{E_k}$  represents the occupation of the  $z$ -direction eigenstate  $|N\rangle$  at in-plane energy  $E_k$ .

When working with periodic systems, it is convenient to work with relative coordinates:  $f_{N,M}^{E_k} \equiv \rho_{N,N+M}^{E_k}$ , with  $M = 0$  terms giving occupations and terms with  $|M| > 0$  represent coherences between states  $|N\rangle$  and  $|N+M\rangle$ . Relative coordinates are convenient because  $f_{N,M}^{E_k}$  is periodic in  $N$ , with a period of  $N_s$ , where  $N_s$  is the number of relevant eigenstates in a single period. We also have  $f_{N,M}^{E_k} \rightarrow 0$  as  $|M| \rightarrow \infty$ , providing an obvious truncation scheme  $|M| \leq N_c$ , where  $N_c$  is a number we refer to as the coherence cutoff. Henceforth, we will assume that  $f$  is simply  $\rho$ , conveniently written in relative coordinates; the relative (absolute) coordinates will be denoted by capital (lowercase) indices.

In the interaction picture (with  $\hbar = 1$ ), the equation of motion for the density matrix can be written as

$$\frac{\partial \tilde{\rho}}{\partial t} = -i[\tilde{H}_i(t), \tilde{\rho}(t)], \quad (2a)$$

$$\tilde{\rho}(t) = \tilde{\rho}(0) - i \int_0^t [\tilde{H}_i(s), \tilde{\rho}(s)] ds, \quad (2b)$$

where  $\tilde{A}(t) = e^{i(H_0+H_{\text{ph}})t} A e^{-i(H_0+H_{\text{ph}})t}$  denotes operator  $A$  in the interaction picture. We will assume that the electron–phonon interaction is weak enough to warrant the assumption that the phonon distribution is negligibly affected by the electrons and write the density matrix as a product  $\tilde{\rho}(t) = \tilde{\rho}_e(t) \otimes \tilde{\rho}_{\text{ph}}$ , where  $\tilde{\rho}_e(t)$  is the electron density matrix and  $\tilde{\rho}_{\text{ph}}$  is the thermal equilibrium density matrix for the phonons. This approximation is called the *Born approximation*. Inserting Eq. (2a) into Eq. (2b) and tracing over the phonon degrees of freedom gives

$$\begin{aligned}\frac{\partial \tilde{\rho}_e}{\partial t} &= -i \text{Tr}_{\text{ph}} \{ [\tilde{H}_i(t), \tilde{\rho}_e(0) \otimes \tilde{\rho}_{\text{ph}}] \} \\ &\quad - \int_0^t ds \text{Tr}_{\text{ph}} \{ [\tilde{H}_i(t), [\tilde{H}_i(s), \tilde{\rho}_e(s) \otimes \tilde{\rho}_{\text{ph}}]] \}.\end{aligned}\quad (3)$$

The first term on the RHS is zero for interaction Hamiltonians that are linear in creation/annihilation operators [45], which is the case for the electron–phonon interaction Hamiltonians considered in this work. An alternative justification for its omission is that we can assume that, after a sufficiently long time, the density matrix has no memory of its initial state  $\tilde{\rho}_e(0)$ . Next, we assume that the time evolution of the density matrix is memoryless (Markov approximation), and replace  $\tilde{\rho}_e(s)$  in the integrand of Eq. (3) by the density matrix at the current time  $\tilde{\rho}_e(t)$ , giving

$$\frac{\partial \tilde{\rho}_e}{\partial t} = - \int_0^t ds \text{Tr}_{\text{ph}} \{ [\tilde{H}_i(t), [\tilde{H}_i(s), \tilde{\rho}_e(t) \otimes \tilde{\rho}_{\text{ph}}]] \}.\quad (4)$$

Equation (4) is the Redfield equation and is not a completely positive map (therefore the density matrix is not guaranteed to remain positive during evolution). Making a change of variables  $s \rightarrow t - s$  and extending the range of integration to infinity (in doing so, we assume environmental correlations decay fast with increasing  $|t - s|$ ) gives

$$\frac{\partial \tilde{\rho}_e}{\partial t} = - \int_0^\infty ds \text{Tr}_{\text{ph}} \{ [\tilde{H}_i(t), [\tilde{H}_i(t - s), \tilde{\rho}_e(t) \otimes \tilde{\rho}_{\text{ph}}]] \}.\quad (5)$$

The net energy change in the electron system due to the two  $\tilde{H}_i$  is zero [44]. The energy change has two components: The in-plane energy change and the transport-direction energy change. Since we assumed that the density matrix is translationally invariant in the in-plane direction, i.e.,  $\langle n\mathbf{k} | \rho | m\mathbf{k}' \rangle = \rho_{n,m}^{E_k} \delta(\mathbf{k} - \mathbf{k}')$ , we will see [Eq. (79) after integration of  $\mathbf{k}_2, \mathbf{k}_3$  and  $\mathbf{k}_4$ ] that the in-plane energy change is zero. Also, because in the transport direction the energy levels are discrete, in the limit  $t \rightarrow \infty$  we can use the rotating-wave approximation, which states the terms resulting in nonzero energy changes average out to zero, and shows that the transport-direction energy change vanishes as well (see Eq. 3.132 in Ref. [44]). Therefore, it will be possible to write Eq. (5) in the Lindblad form, i.e., the equation is completely

positive and trace preserving. We remove the remaining time dependence by switching back to the Schrödinger picture:

$$\begin{aligned}\frac{\partial \rho_e}{\partial t} &= -i[H_0, \rho_e(t)] - \int_0^\infty ds \text{Tr}_{\text{ph}} \{ [H_i, [e^{-i(H_0+H_{\text{ph}})s} H_i e^{i(H_0+H_{\text{ph}})s}, \rho_e(t) \otimes \rho_{\text{ph}}]]\},\end{aligned}\quad (6)$$

The above term can be written as a sum of unitary time evolution and a dissipative term that includes interaction with phonons,

$$\frac{\partial \rho_e}{\partial t} = -i[H_0, \rho_e(t)] + D[\rho_e(t)],\quad (7)$$

where the superoperator  $D$  describes dissipation and is often referred to as the dissipator. “Appendix 2” is devoted to the derivation of the dissipator for electron–phonon interaction while interactions with other scattering mechanisms are covered in “Appendix 3”. Again, the dissipator was derived according to a procedure that guarantees the Lindblad form (see more in [42, 44, 82, 83]). However, once can readily convince oneself that this is true for any particular mechanism, as shown in “Appendix 4” on the example of electron–phonon interaction.

Equation (7) written for the matrix elements  $f_{N,M}^{E_k}$  in relative coordinate has the form

$$\frac{\partial f_{N,M}^{E_k}}{\partial t} = -i \frac{\Delta_{N,N+M}}{\hbar} f_{N,M}^{E_k} + \mathcal{D}_{N,M}^{E_k},\quad (8)$$

where  $\Delta_{N,N+M}$  is the energy spacing between states  $N$  and  $M$ . For brevity, we use  $\mathcal{D}$  to denote  $Df$  and will henceforth call this whole term (rather than the superoperator) the dissipator.

$$\begin{aligned}\mathcal{D}_{N,M}^{E_k} &= - \sum_{n,m,g,\pm} I_{NMnmE_k}^{\text{out},g,\pm} f_{N,n}^{E_k} \\ &\quad + \sum_{n,m,g,\pm} I_{NMnmE_k}^{\text{in},g,\pm} f_{N+n,M+m-n}^{E_k \mp E_g + \Delta_{N+M,N+M+m}} + \text{h.c.}\end{aligned}\quad (9)$$

Here,  $I_{NMnmE_k}^{\text{in},g,\pm}$  and  $I_{NMnmE_k}^{\text{out},g,\pm}$  are the quantum-mechanical generalizations of semiclassical scattering rates [9, 16]. They are given in “Appendix 5”. In this context, h.c. should be understood as the matrix elements of the Hermitian conjugate (it is not simply the complex conjugate because indices need to be swapped as well). In this work, we will refer to them as scattering rates (or simply rates). The  $g$  index represents the different scattering mechanisms,  $+(-)$  refers to absorption (emission), and  $E_g$  is the energy associated with scattering mechanism  $g$  (i.e., phonon energy). For elastic scattering mechanisms, such as ionized-impurity scattering, we have  $E_g = 0$ .

In this work, we are interested in the steady-state solution of Eq. (8). Once the steady-state density matrix is known,

experimentally measurable quantities can easily be calculated. Detailed information regarding the different observables, such as the current density and charge density, is given in Sect. 2.1. In Sect. 2.2, we derive an expression for the optical gain by considering the linear response of the density matrix to an optical field, treating it as a small perturbation. This section concludes with Sect. 2.3, where we provide a comparison with semiclassical model based on the Pauli master equation [46], where only the diagonal values of the density matrix are included.

## 2.1 Steady-state observables

Once the steady-state density-matrix elements have been calculated from Eq. (8), the expectation value of any observable  $A$  can be calculated using

$$\langle A \rangle = \sum_{N=1}^{N_s} \sum_{M=-N_c}^{N_c} \int dE_k \langle N|A|N+M \rangle f_{N,M}^{E_k}, \quad (10)$$

Here,  $|N\rangle$  represents an eigenstate of  $H_0$  in Dirac notation. As before,  $N_s$  is the number of eigenstates per period and  $N_c$  is a coherence cutoff, denoting how much coupling there is between the states in the neighboring stages (see further discussion in Sect. 3.1). The observable  $A$  can depend on position  $z$ , any derivative of position  $\frac{\partial^n}{\partial z^n}$ , as well as in-plane energy. An important steady-state observable is the drift velocity  $v$ , which is used to calculate the expectation value of current density  $J = qn_{3D}\langle v \rangle$ , where  $n_{3D}$  is the average 3D electron density in the device,

$$n_{3D} = \frac{1}{L_p} \int_0^{L_p} N_D(z) dz, \quad (11)$$

with  $L_p$  the period length of the QCL and  $N_D(z)$  the doping density. Matrix elements of the drift velocity operator are obtained from the time derivative of the position operator using

$$\begin{aligned} \langle n|v|m \rangle &= \left\langle n \left| \frac{dz}{dt} \right| m \right\rangle = \frac{i}{\hbar} \langle n|[H, z]|m \rangle \\ &= \frac{i}{\hbar} (E_n - E_m) \langle n|z|m \rangle = \frac{i}{\hbar} (E_n - E_m) z_{n,m}, \end{aligned} \quad (12)$$

where we have used  $\langle n|[H_i, z]|m \rangle = 0$  because the interaction potentials in this work only contain the position operator  $z$ , but not momentum  $p_z$ . The steady-state current density can be written as

$$J = qn_{3D} \sum_{N=1}^{N_s} \sum_{M=-N_c}^{N_c} \int dE_k v_{N,N+M} f_{N,M}^{E_k}. \quad (13)$$

In this work, electron-electron interaction is treated at a mean-field level by obtaining a self-consistent solution with

Poisson's equation (see further discussion in Sect. 3.1). We therefore need to calculate the spatially resolved electron density  $n_{el}(z)$  using

$$\begin{aligned} n_{el}(z) &= n_{3D} \sum_{N=1}^{N_s} \sum_{M=-N_c}^{N_c} \int dE_k \\ &\times \sum_b \left[ \psi_N^{(b)}(z) \right]^* \psi_{N+M}^{(b)}(z) f_{N,M}^{E_k}, \end{aligned} \quad (14)$$

Here,  $\psi_N^{(b)}(z)$  is the  $N$ th electron eigenfunction in band  $b$ . The sum involving  $b$  is over the three-bands considered in the  $\mathbf{k} \cdot \mathbf{p}$  model described in “Appendix 1”. We define the occupation of state  $N$  using

$$f_N = \int dE_k f_{N,0}^{E_k} \quad (15)$$

and the average coherence between states  $N$  and  $M$  as

$$\rho_{N,M} = \int dE_k \rho_{N,M}^{E_k} = \int dE_k f_{N,M-N}^{E_k}. \quad (16)$$

The magnitude of the coherences relative to the occupations give us an idea how important is it to include off-diagonal elements of the density matrix and when a semiclassical treatment is sufficient.

In order to quantify the amount of electron heating that takes place under operating conditions, we define the expectation value of in-plane energy using

$$\langle E_k \rangle = \sum_{N=1}^{N_s} \int dE_k E_k f_{N,0}^{E_k} \quad (17)$$

and the expectation value of in-plane energy of a single subband using

$$\langle E_k \rangle_N = \frac{\int dE_k E_k f_{N,0}^{E_k}}{\int dE_k f_{N,0}^{E_k}}. \quad (18)$$

For low bias, we have  $\langle E_k \rangle_N \simeq k_B T$ , where  $T$  is the lattice temperature. However, as will be shown in Sect. 4.1.4, for electric fields around and above threshold,  $\langle E_k \rangle_N$  can greatly exceed  $k_B T$  at lattice temperature and can differ substantially between subbands. It is possible to define the electron temperature of state  $N$  using  $T_N = \langle E_k \rangle_N / k_B$ . The electron temperature can be a useful quantity because it is an input parameter for simplified DM models, where a thermal in-plane distribution is assumed [22, 34]. However, we note that the in-plane energy distribution under typical operational electric fields can deviate strongly from a Maxwellian. The term electron temperature should therefore only be interpreted as the temperature equivalent of the average in-plane energy.

## 2.2 Optical gain

The  $z$ -component of an electric field propagating in the  $x$ -direction, with frequency  $\omega$  and amplitude  $\mathcal{E}_0$ , can be written as a plane wave of the form [47]

$$\mathcal{E}_z(x, \omega) = \mathcal{E}_0 e^{in_B \omega x/c} e^{-\alpha(\omega)x/2}, \quad (19)$$

where  $c$  is the speed of light in vacuum,  $n_B$  is the background index of refraction (which we assume is frequency independent in the frequency range of interest in this work), and  $\alpha(\omega)$  is the absorption coefficient. We note that the electric field in typical QCL waveguides (operating in a TM mode) also has a nonzero  $x$ -component [9]. However, only the dipole matrix element in the  $z$  direction are nonzero, so the  $x$ -component is not amplified. Optical gain is defined as  $g(\omega) = -\alpha(\omega)$  and can be obtained from the complex susceptibility using [47]

$$g(\omega) = -\frac{\omega}{cn_B} \text{Im}[\chi(\omega)]. \quad (20)$$

In order to calculate the complex susceptibility, we follow the approach in Ref. [31]. We treat the optical field as a small perturbation and consider the linear response in the frequency domain. We assume the wavelength of the optical field is long and ignore the position dependence of the electric field across the QCL. We can then write the perturbation potential as

$$V_\omega(z) = q\mathcal{E}_0 z e^{-i\omega t}. \quad (21)$$

Using linear-response theory, the density matrix can be written as

$$\rho(t) = \rho_0 + \rho_\omega e^{-i\omega t}, \quad (22)$$

where  $\rho_0$  is the steady-state solution for  $\mathcal{E}_0 = 0$ . The equation of motion for  $\rho(t)$  can be written as

$$\begin{aligned} i\hbar \frac{\partial \rho}{\partial t} &= [H_0 + V_\omega, \rho_0 + \rho_\omega] + \mathcal{D}[\rho_0 + \rho_\omega] \\ &= [H_0, \rho_0] + \mathcal{D}[\rho_0] + [H_\omega, \rho_\omega] + [H_0, \rho_\omega] \\ &\quad + [V_\omega, \rho_0] + \mathcal{D}[\rho_\omega] \\ &\simeq [H_0, \rho_\omega] + [V_\omega, \rho_0], \end{aligned} \quad (23)$$

where the first two terms in the second line of Eq. (23) add up to zero and the third term is proportional to  $e^{-2i\omega t}$  (second harmonic). The second-harmonic term is a product of two small quantities ( $V_\omega$  and  $\rho_\omega$ ) and is ignored. Combining Eqs. (21)–(23) and dividing both sides by  $e^{-i\omega t}$  gives an equation for  $\rho_\omega$

$$\hbar\omega\rho_\omega = [H_0, \rho_\omega] + \mathcal{D}[\rho_\omega] + e\mathcal{E}_0[z, \rho_0]. \quad (24)$$

In the same way, we defined the matrix elements  $f_{NM}^{E_k}$ , we can define the matrix elements of  $\rho_\omega$  in relative coordinates

as  $f_{NM}^{E_k}(\omega) = \rho_{N,N+M}^{E_k}(\omega)$ . Equation (24) gives the following equation for the matrix elements

$$f_{N,M}^{E_k}(\omega) \left[ i \frac{A_{N,N+M}}{\hbar} - i\omega \right] = A_{N,M}^{E_k} + \mathcal{D}_{N,M}^{E_k}(\omega), \quad (25)$$

where  $\mathcal{D}_{N,M}^{E_k}(\omega)$  are matrix elements of the dissipator acting on the linear-response density matrix and

$$A_{N,M}^{E_k} = -i \frac{e\mathcal{E}_0}{\hbar} \sum_n z_{N,N+n} f_{N+n,M-n}^{E_k} + \text{h.c.}, \quad (26)$$

where  $z_{n,m}$  is a matrix elements of the  $z$ -position operator. Using Eqs. (25) and (26), we can calculate  $f_{NM}^{E_k}(\omega)$  for relevant values of  $\omega$ . Note that the dissipator term depends on many of the matrix elements  $f_{N,M}^{E_k}(\omega)$ , so solving for  $f_{N,M}^{E_k}(\omega)$  is not as trivial as it might appear from Eq. (25). Our numerical method for solving Eq. (25) is given in Sect. 3.

Once  $f_{NM}^{E_k}(\omega)$  is known, we can calculate the field-induced polarization using

$$p(\omega) = \text{Tr}[d\rho_\omega] = e \int dE_k \sum_{N,M} z_{N,N+M} f_{N,M}^{E_k}(\omega), \quad (27)$$

where  $d$  is the dipole operator  $d = qz$ . Note that the density matrix is normalized to one, so  $p(\omega)$  gives the induced polarization per electron. To get the polarization per unit volume, we use  $\mathcal{P}(\omega) = n_{3D}p(\omega)$ , where  $n_{3D}$  is the average electron density. The polarization is then used to calculate the complex susceptibility  $\chi(\omega)$  using [31, 47]

$$\chi(\omega) = \frac{\mathcal{P}(\omega)}{\epsilon_0 \mathcal{E}_0}. \quad (28)$$

Alternatively, the complex susceptibility can be calculated using the induced current density, which is the time derivative of the polarization  $J(\omega) = -i\omega\mathcal{P}$ , giving

$$\chi(\omega) = \frac{i}{\omega} \frac{J(\omega)}{\epsilon_0 \mathcal{E}_0}. \quad (29)$$

We have calculated the susceptibility using both Eq. (28) and Eq. (29), and both expressions give the same result (within numerical accuracy). Once the complex susceptibility has been calculated, the gain is obtained using Eq. (20).

## 2.3 Comparison with semiclassical models

Equation (8) can be thought of as a generalization of the semiclassical Pauli master equation, wherein the rates are calculated using Fermi's golden rule [9, 16]. This fact can be seen by looking at the diagonal terms ( $M = 0$ ) and limiting the out-scattering sum to  $n = 0$ , and the in-scattering sum to  $n = m$ , giving

$$\begin{aligned} \frac{\partial f_{N,0}^{E_k}}{\partial t} = & -2 \sum_{m,g,\pm} \Gamma_{N,0,0,m,E_k}^{\text{out},g,\pm} f_{N,0}^{E_k} \\ & + 2 \sum_{n,g,\pm} \Gamma_{N,0,n,n,E_k}^{\text{in},g,\pm} f_{N+n,0}^{E_k + \Delta_{N,N+n} \mp E_s}. \end{aligned} \quad (30)$$

By defining

$$\begin{aligned} f_N^{E_k} &= f_{N,0}^{E_k}, \\ \tilde{\Gamma}_{N,n,E_k}^{\text{out},g,\pm} &= 2\Gamma_{N,0,0,n,E_k}^{\text{out},g,\pm}, \\ \tilde{\Gamma}_{N,n,E_k}^{\text{in},g,\pm} &= 2\Gamma_{N,0,n,n,E_k}^{\text{in},g,\pm}, \end{aligned} \quad (31)$$

we can write Eq. (30)

$$\begin{aligned} \frac{\partial f_N^{E_k}}{\partial t} = & - \sum_{n,g,\pm} \tilde{\Gamma}_{N,n,E_k}^{\text{out},g,\pm} f_N^{E_k} \\ & + \sum_{n,g,\pm} \tilde{\Gamma}_{N,n,E_k}^{\text{in},g,\pm} f_{N+n}^{E_k + \Delta_{N,N+n} \mp E_s}, \end{aligned} \quad (32)$$

which is equivalent to the Boltzmann equation for the occupation of state  $N$  with in-plane energy  $E_k$  (see, for example, Eq. (129) in Ref. [9]). The factor of 2 in front of the rates comes from the addition of the Hermitian conjugate (which is identical for the diagonal of the density matrix). Because of the similarities between the semiclassical Boltzmann equation in (32) and the quantum-mechanical results in Eq. (8), we can easily compare results obtained using both models and investigate when semiclassical models give satisfactory results and when off-diagonal density-matrix elements must be included.

When calculating current density using semiclassical models, we cannot directly use Eq. (13) because matrix elements of the velocity operator given in Eq. (12) are zero for  $n = m$  (in other words, all the current is given by off-diagonal matrix elements). In order to calculate drift velocity, we first define

$$\tilde{\Gamma}_{NnE_k}^{\text{out}} = \sum_{g,\pm} \tilde{\Gamma}_{N,n,E_k}^{\text{out},g,\pm}, \quad (33)$$

which gives the total rate at which an electron in subband  $N$  with in-plane energy  $E_k$  is scattered to subband  $N + n$ . We then define the semiclassical drift velocity as

$$v_c = \int dE_k \sum_{N=1}^{N_s} \sum_{n=-N_c}^{N_c} (z_N - z_{N+n}) f_N^{E_k} \Gamma_{NnE_k}^{\text{out}}, \quad (34)$$

with  $z_N = \langle \psi_N | z | \psi_N \rangle$ . The current density is then calculated using  $J_c = qn_{3D}v_c$ . Equation (34) can be interpreted as summing up the flux of electrons from position  $z_N$  to  $z_{N+n}$ . This intuitive expression for  $v_c$  can be derived from Eq. (13) by approximating the off-diagonal density-matrix elements using diagonal ones [48].

If the optical frequency  $\omega$  is much larger than typical scattering rates (due to phonons, impurities etc.), the gain can be calculated from the diagonal density-matrix elements using [9, 49, 50]

$$\begin{aligned} g(\omega) = & \frac{q^2 \omega m_{||}^*}{\hbar^3 c \epsilon_0 n_B L_P} \sum_{N=1}^{N_s} \sum_{M=-N_c}^{N_c} \frac{\Delta_{N,N+M}}{|\Delta_{N,N+M}|} |z_{N,N+M}|^2 \\ & \times \int dE_k (f_N^{E_k} - f_{N+M}^{E_k}) \mathcal{L}_{N,N+M}(E_k, \omega), \end{aligned} \quad (35)$$

with the Lorentzian broadening function

$$\mathcal{L}_{n,m}(E_k, \omega) = \frac{\gamma_{n,m}(E_k)}{\gamma_{n,m}^2(E_k) + [\omega - \Delta_{nm}/\hbar]^2}. \quad (36)$$

The broadening of a transition  $\gamma_{n,m}(E_K)$  is calculated using [9]

$$\gamma_{n,m}(E_k) = \frac{1}{2} [\gamma_n(E_k) + \gamma_m(E_k)], \quad (37)$$

where  $\gamma_n(E_k)$  is the total intersubband scattering rate from subband  $n$  at energy  $E_k$ , calculated using

$$\gamma_n(E_k) = \sum_{n \neq 0} \tilde{\Gamma}_{NnE_k}^{\text{out}}. \quad (38)$$

### 3 Numerical method

#### 3.1 Steady-state density matrix

The central quantities of interest are the matrix elements of the steady-state density matrix  $f_{N,M}^{E_k} = \rho_{N,N+M}^{E_k}$ . Because of periodicity, we have  $f_{N,M}^{E_k} = f_{N+N_s,M}^{E_k}$ , where  $N_s$  is the number of states in a single period. The index  $M$  strictly runs from  $-\infty$  to  $+\infty$  but is truncated according to  $|M| \leq N_c$ , where  $N_c$  is a coherence cutoff. The coherence cutoff needs to be large enough such that overlaps between states  $|N\rangle$  and  $|N + N_c\rangle$  is small for all  $N$ . Typical values for  $N_c$  are slightly larger than  $N_s$ ; however, the specific choice of  $N_c$  is highly system dependent. We also have to discretize the continuous in-plane energy  $E_k$  into  $N_E$  evenly spaced values. This truncation scheme results in  $N_s(2N_c + 1)N_E$  matrix elements.

In order to calculate the steady-state matrix elements  $f_{N,M}^{E_k}$ , we set the time derivative in Eq. (8) to zero and get

$$\begin{aligned} 0 = & -i \frac{\Delta_{N,N+M}}{\hbar} f_{N,M}^{E_k} - \sum_{n,m,g,\pm} \Gamma_{NMnmE_k}^{\text{out},g,\pm} f_{N,n}^{E_k} \\ & + \sum_{n,m,g,\pm} \Gamma_{NMnmE_k}^{\text{in},g,\pm} f_{N+n,M+m-n}^{E_k \mp E_0 + \Delta_{N+M,N+M,m}} + \text{h.c..} \end{aligned} \quad (39)$$

Note that the sums in the Eq. (39) run over matrix elements and energies outside the center period (e.g.,  $N > N_s$  or  $N < 1$ ), which are calculated by taking advantage of periodicity. We also encounter terms for which  $|M| > N_c$ , where we assume  $f_{N,M}^{E_k} = 0$ . The exact values of  $N_s$ ,  $N_c$ , and  $N_E$  are very system specific, for example, in order to simulate the 8.5 μm-QCL considered in Sect. 4.1, we use  $N_s = 8$ ,  $N_c = 12$  and  $N_E = 251$  resulting in 50, 200 matrix elements and for the 4.6-μm-QCL studied in Sect. 4.2 we use  $N_s = 12$ ,  $N_c = 14$ , and  $N_E = 251$  resulting in 87, 348 matrix elements. This high number of variables makes it impractical to solve Eq. (39) as a matrix equation, so we resorted to iterative methods.

In order to solve Eq. (39) using the Jacobi iterative method, [51] we need to write  $f_{N,M}^{E_k}$  in terms of all other terms  $f_{N',M'}^{E'_k}$  for  $N' \neq N$ ,  $M' \neq M$ , and  $E'_k \neq E_k$ . From the dissipator term in (39), we see that the terms containing  $f_{N,M}^{E_k}$  in the out-scattering part are the  $n = M$  terms. As for the in-scattering part, only the elastic terms ( $E_0 = 0$ ) with  $n = m = 0$  contain  $f_{N,M}^{E_k}$ . Note that the Hermitian conjugate part also contains terms proportional to  $f_{N,M}^{E_k}$  which are obtained using  $f_{N,M}^{E_k} = [f_{N+M,-M}^{E_k}]^*$ , which is equivalent to  $\rho_{N,M}^{E_k} = [\rho_{M,N}^{E_k}]^*$ . It is convenient to define the reduced dissipator  $\tilde{\mathcal{D}}_{N,M}^{E_k}$ , where the terms containing  $f_{N,M}^{E_k}$  are excluded. Summing up all the terms multiplying  $f_{N,M}^{E_k}$  gives

$$\begin{aligned} \gamma_{NME_k} = & - \sum_{m,g,\pm} \left[ \Gamma_{N,M,M,m,E_k}^{\text{out},g\pm} + \Gamma_{N+M,-M,m,E_k}^{\text{out},g\pm} \right] \\ & + \sum_{E_0=0,\pm} \left[ \Gamma_{N,M,0,0,E_k}^{\text{in},g,\pm} + \Gamma_{N+M,-M,0,0,E_k}^{\text{in},g,\pm} \right], \end{aligned} \quad (40)$$

where  $E_0 = 0$  represents summing over all elastic scattering mechanisms. Using Eq. (40), we can write the reduced dissipator as

$$\tilde{\mathcal{D}}_{N,M}^{E_k} = \mathcal{D}_{N,M}^{E_k} - \gamma_{NME_k} f_{NM}^{E_k}, \quad (41)$$

which can depend on all matrix elements of the density matrix except for  $f_{NM}^{E_k}$ . Using Eqs. (41)–(40), we can write Eq. (39) as

$$f_{N,M}^{E_k} = \frac{\tilde{\mathcal{D}}_{N,M}^{E_k}}{\frac{i}{\hbar} \Delta_{N,N+M} - \gamma_{N,M,E_k}}, \quad (42)$$

where the RHS does not depend on  $f_{N,M}^{E_k}$ . Equation (42) is then used as a basis for an iterative solution for  $f_{NM}^{E_k}$  using a weighted Jacobi scheme [51]

$$f_{NM}^{E_k,(j)} = (1 - \xi) f_{NM}^{E_k,(j-1)} \quad (43)$$

$$+ \xi \frac{\mathcal{D}_{NM}^{E_k,(j-1)} - \gamma_{NME_k} f_{NM}^{E_k,(j-1)}}{\frac{i}{\hbar} \Delta_{N,N+M} - \gamma_{N,M,E_k}}, \quad (44)$$

where  $\xi \in [0, 1]$  is a weight. The quantity  $\mathcal{D}_{NM}^{E_k,(j)}$  should be understood as matrix elements of the dissipator acting on the  $j$ th iteration of the density matrix. As an initial guess, we assume a diagonal density matrix, where all subbands are equally occupied with a thermal distribution,

$$f_{NM}^{E_k,(1)} = A \delta_{M,0} e^{-E_k/(k_B T)}, \quad (45)$$

where  $A$  is normalization constant. In the current work, we used  $\xi = 0.6$ . However, we did not find convergence speed to depend strongly on the specific value. However, we found the numerical method to be more stable when  $\xi > 0.5$ . We iterate Eq. (43) until the following convergence criterion is satisfied:

$$\sum_{N,M} \int dE_k |f_{NM}^{E_k,(j)} - f_{NM}^{E_k,(j-1)}| < \delta_{\text{iter}}, \quad (46)$$

where we used  $\delta_{\text{iter}} = 10^{-3}$  in this work. Note that the density matrix is normalized to one, making  $\delta_{\text{iter}}$  unitless. Typically, 50–150 iterations are needed to reach the convergence criteria.

In order to include corrections due to electron–electron interaction on a mean-field level, the iterative solution method described above is supplemented with a self-consistent solution to Poisson's equation. The solutions to Poisson's equation are obtained using an outer self-consistency loop (the inner loop is the iterative solution described earlier). At the end of the inner (iterative) loop, corrections to the band structure  $V_p(z)$  are calculated based on a self-consistent solution with Poisson's equation. If the change in  $V_p(z)$  is above a certain threshold, the eigenfunctions and rates are recalculated and the inner iterative loop is repeated with the new eigenfunctions and rates. The process is repeated until  $V_p(z)$  converges. To quantify when convergence is reached, we define the largest first-order correction to the eigenfunctions

$$\delta_{\text{self}} = \max_m \left| \int dz |\psi_m(z)|^2 [V_p^{\text{new}}(z) - V_p^{\text{old}}(z)] \right|, \quad (47)$$

where  $|\psi_m(z)|^2$  is the probability density of eigenfunction  $m$ . In this work, we use  $\delta_{\text{self}} = 0.1 \text{ meV}$ , which is small compared with a typical energy spacing in QCLs. The number of iterations needed to satisfy the convergence criterion in Eq. (47) depends on the electron density in the device (higher density means more iterations). For the device considered in Sect. 4.1, 2–5, iterations are needed for the first value of applied electric field. For other values of electric field, the number of iteration needed can be kept

to a minimum by sweeping the field and using results from previous fields as the initial guess for the next value of the electric field.

### 3.2 Linear-response density matrix

Once we have calculated the steady-state density matrix using the procedure described in Sect. 3.1, we can calculate the linear-response density-matrix elements  $f_{NM}^{E_k}(\omega)$  using Eqs. (25) and (26). By following the same steps as in Sect. 3.1, we get

$$f_{NM}^{E_k(j)}(\omega) = (1 - \xi)f_{NM}^{E_k(j-1)}(\omega) + \xi \frac{\mathcal{D}_{NM}^{E_k(j-1)}(\omega) - \gamma_{NME_k} f_{NM}^{E_k(j-1)}(\omega) + A_{N,M}^{E_k(j-1)}}{\frac{i}{\hbar}(\Delta_{N,N+M} - \hbar\omega) - \gamma_{NME_k}}. \quad (48)$$

where the matrix elements  $A_{N,M}^{E_k(j)}$  only depend on the steady-state density matrix and are given by Eq. (26). Note that the matrix elements  $\mathcal{D}_{NM}^{E_k(j)}(\omega)$  cannot be calculated directly using Eq. (9) because Fourier components of the linear-response density matrix are not Hermitian and the resulting dissipator cannot be written as a pair of Hermitian conjugate terms. The calculations will not be shown here, but the result is

$$\begin{aligned} \mathcal{D}_{NM}^{E_k}(\omega) = & - \sum_{n,m,g,\pm} \Gamma_{NMnmE_k}^{\text{out},g\pm} f_{N,n}^{E_k}(\omega) \\ & + \sum_{n,m,g,\pm} \Gamma_{NMnmE_k}^{\text{in},g,\pm} f_{N+n,M+m-n}^{E_k \mp E_s + \Delta_{N+M,N+M+m}}(\omega) \\ & - \sum_{n,m,g,\pm} \Gamma_{N+M,-M,n,m,E_k}^{\text{out},g\pm} f_{N+M+n,-n}^{E_k}(\omega) \\ & + \sum_{n,m,g,\pm} \Gamma_{N+M,-M,n,m,E_k}^{\text{in},g,\pm} f_{N+m,M-m+n}^{E_k \mp E_s + \Delta_{N,N+m}}(\omega), \end{aligned} \quad (49)$$

which would reduce to Eq. (9) if the linear response DM were Hermitian. Equation (48) is iterated until

$$|g^{(j)}(\omega) - g^{(j-1)}(\omega)| < 10^{-3} \text{ cm}^{-1}, \quad (50)$$

where  $g^{(j)}(\omega)$  is the gain calculated from  $f_{NM}^{E_k(j)}(\omega)$ , using Eq. (28). This process is repeated for multiple values of  $\omega$  in the frequency range of interest.

## 4 Results

### 4.1 Lattice-matched design

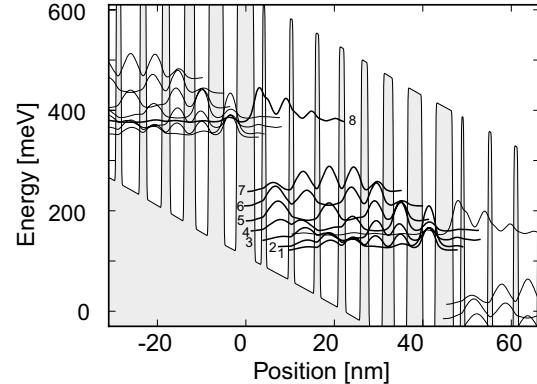
In order to verify the accuracy of our model, we simulated a mid-infrared QCL proposed in Ref. [52]. We chose this device because it has previously been successfully modeled using both NEGF and simplified density-matrix approaches [34], which allowed us to compare our results to existing theoretical results. The QCL was designed for wide voltage

tuning, emitting around  $8.5 \mu\text{m}$  (146 meV), with a tuning range of almost  $100 \text{ cm}^{-1}$  (12 meV). The device is based on the  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}/\text{In}_{0.52}\text{Al}_{0.48}\text{As}$  material system, which is lattice-matched to InP, eliminating the effects of strain. With this material composition,  $\text{In}_{0.47}\text{Ga}_{0.53}\text{As}$  acts as potential wells and  $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$  as barriers with approximately a 520 meV conduction-band discontinuity between materials. All material parameters relating to band structure (e.g., Kane energy) as well as other constants such as phonon energy and interface-roughness parameters are provided in “Appendix 6”.

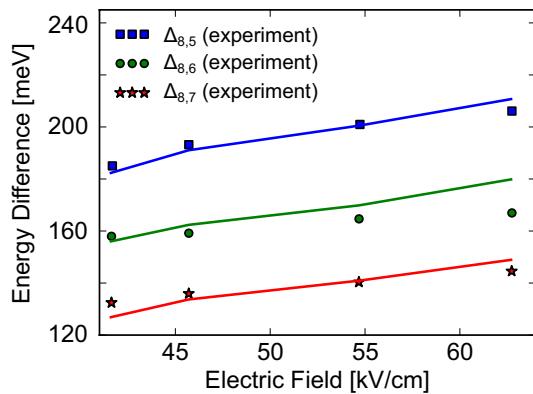
#### 4.1.1 Band structure

Figure 1 shows the conduction-band diagram and the probability density of the relevant eigenfunctions slightly above threshold (50 kV/cm). The figure shows the 8 considered states belonging to a single period in bold and states belonging to neighboring periods are represented as thin curves. The lasing transition is between states 8 and 7, and state 6 is the main extractor state. State number 1 is the energetically lowest state in the injector (often denoted as the ground state).

In order to verify the accuracy of the band structure shown in Fig. 1, we compare various calculated energy differences  $\Delta_{n,m}$  to results obtained from experimentally determined electroluminescence spectra [52]. Figure 2 shows the calculated energy differences  $\Delta_{8,7}$ ,  $\Delta_{8,6}$  and  $\Delta_{8,5}$  as a function of electric field (symbols) and a comparison



**Fig. 1** Conduction-band edge (thin curve bounding gray area) and probability densities for the 8 eigenstates used in calculations (bold curves) for a field strength of 50 kV/cm. States belonging to neighboring periods are denoted by thin gray curves. The states are numbered in increasing order of energy, starting with the ground state in the injector. Using this convention, the lasing transition is from state 8 to state 7. The length of one period is 44.9 nm, with a layer structure (in nm), starting with the injector barrier (centered at the origin) **4.0/1.8/0.8/5.3/1.0/4.8/1.1/4.3/1.4/3.6/1.7/3.3/2.4/3.1/3.4/2.9**, with barriers denoted bold. Underlined layers are doped to  $1.2 \times 10^{17} \text{ cm}^{-3}$ , resulting in an average charge density of  $n_{3D} = 1.74 \times 10^{16} \text{ cm}^{-3}$  over the stage

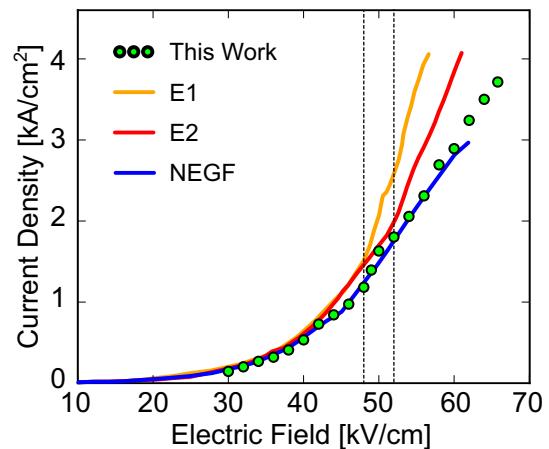


**Fig. 2** Calculated energy differences  $\Delta_{8,7}$  (lowest red curve),  $\Delta_{8,6}$  (middle green curve) and  $\Delta_{8,5}$  (top blue curve) as a function of field. The symbols show results based on experimentally determined electroluminescence spectra from Ref. [52] (Color figure online)

to experimentally obtained values (curves). From Fig. 2, we can see that for low field strength (41.6 and 45.9 kV/cm), the deviation from experiment is small and the largest error is 3 meV (less than a 2% relative difference) for the  $\Delta_{8,6}$  energy spacing for a field strength of 46 kV/cm. For higher field strengths (54.7 and 62.7 kV/cm), the energy differences are overestimated by up to 13 meV (8% relative error) for the  $\Delta_{8,6}$  energy spacing for a field strength of 62.7 kV/cm. The reason for the discrepancy between experiment and theory at high fields is not clear. A possible reason is the low bandgap of the well material ( $\sim 0.74$  meV) compared with the lasing transition at higher fields ( $\sim 0.15$  eV), limiting the accuracy of the three-band  $\mathbf{k} \cdot \mathbf{p}$  model.

#### 4.1.2 Current density versus field

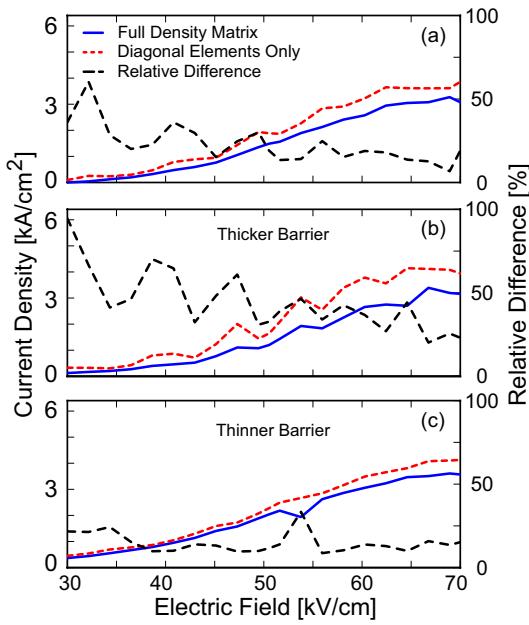
Figure 3 shows results for current density vs electric field calculated using the density-matrix formalism presented in this work, along with a comparison with two different experiments and theoretical results based on the NEGF formalism. Note that neither theoretical result takes the laser field into account, so we can only expect results to agree with experiment below and around threshold (denoted by dashed vertical lines in Fig. 3). The two experimental results are based on the same QCL-core design but with different waveguide designs. Experiment 1 (E1) refers to the sample from Ref. [52], which is based on a buried-heterostructure waveguide design. Experiment 2 (E2) refers to a sample based on a double-trench waveguide design with higher estimated losses than experiment 1 [34]. From Fig. 3, we can see that our results based on the density-matrix formalism shows excellent agreement with the NEGF results for low fields (less than 40 kV/cm) and slightly higher current density for higher fields. Our density-matrix results are in quantitative agreement with experiments up to threshold for



**Fig. 3** Calculated current density vs field based on this work and comparison to results using NEGF [34]. Also shown are experimental results from Ref. [52] (E1) and a regrown device from Ref. [34] (E2). Experimentally determined threshold fields of 48 kV/cm (E1) and 52 kV/cm (E2) are denoted by dashed vertical lines. Both theoretical results are given for a lattice temperature of 300 K. Experimental results are provided for pulsed operation, minimizing the effect of lattice heating above the heatsink temperature of 300 K (Color figure online)

both experiments, while the NEGF underestimate current density for fields higher than 40 kV/cm. Above threshold, our density-matrix results underestimate the current density due to the omission of the laser electromagnetic field.

In order to get an idea how important the inclusion of off-diagonal density-matrix elements are, Fig. 4 shows a comparison between the current density obtained using the full density matrix and results obtained using only the diagonal matrix elements (semiclassical results). The importance of including off-diagonal matrix elements is known to be greater for thicker injection barriers, where transport is limited by tunneling through the injection barrier [39]. For this reason, results are also plotted for different injection barrier widths in Fig. 4b, c. Figure 4a shows the results for the same device as in Fig. 3 (4-nm-thick injection barrier), while (b) and (c) show the results for a thicker (5 nm) and thinner (3 nm) injection barrier, respectively (but otherwise identical devices). In Fig. 4a, we see that even for a mid-infrared QCL, where transport has previously been reported to be mostly incoherent, [14] the difference is noticeable and the current density is overestimated by about 10–60% by semiclassical methods. The relative difference around threshold ( $\sim 48$  kV/cm) is about 29%. This is a much bigger difference than the previously reported value of only a few percent for a different mid-infrared QCL design [14]. The difference could be a result of the different device design, or the assumption of a factorized (in-plane and cross-plane) density matrix employed in Ref. [14], while we do not make that assumption. Figure 4b shows that, for a thicker injection

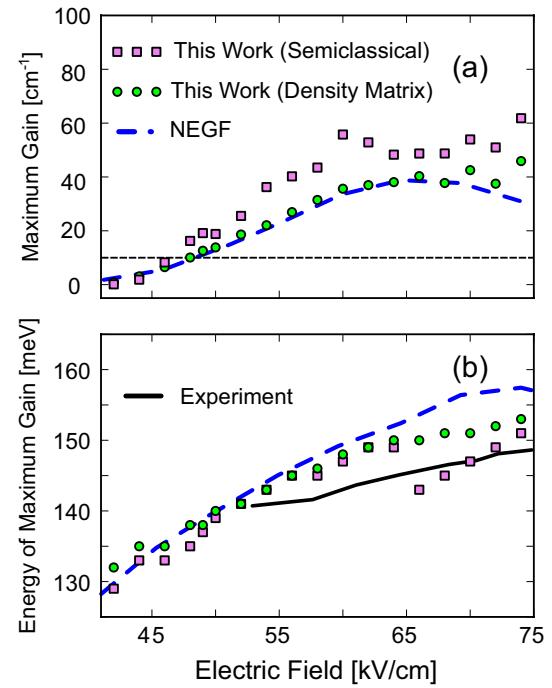


**Fig. 4** Comparison between the current density obtained using the full density matrix (solid blue curve) and semiclassical results with only diagonal elements included (short-dashed red curve). The relative difference between the two results is also shown (long-dashed black curve). **a** Results for the same device considered in Fig. 3. **b** Results for a device with a thicker injection barrier (5 nm). **c** Results for a device with a thinner injection barrier (3 nm) (Color figure online)

barrier (5 nm), a semiclassical model fails completely, with a relative difference ranging between 22 and 90% and about 30% difference around threshold. The semiclassical results also show peaks and negative differential resistance (NDR) regions that are absent (or much less pronounced) in the DM results. Figure 4c shows that, for the thinner barrier (3 nm), the difference is smaller (between 8 and 32%) and about 9% around threshold. We can see that the considered device design with a 4-nm-thick injection barrier is on the border of admitting a semiclassical description.

#### 4.1.3 Optical gain and threshold current density

Figure 5a shows the peak gain as a function of electric field calculated using our density-matrix and semiclassical models. Also shown is a comparison with the results based on NEGF [34]. The estimated threshold gain [34, 52]  $G_{\text{th}} \simeq 10 \text{ cm}^{-1}$  from experiment E1 is denoted by a horizontal dashed line. From Fig. 5a, we see that our DM results predict a threshold field of  $E_{\text{th}} = 48.0 \text{ kV/cm}$ . By matching this field to Fig. 3, a threshold current density of  $J_{\text{th}} = 1.55 \text{ kA/cm}^2$  is found. The results are in good agreement with experiment ( $E_{\text{th}} = 48.0 \text{ kV/cm}$  and  $J_{\text{th}} = 1.50 \text{ kA/cm}^2$ ). Our results are also in fair agreement with the NEGF results ( $E_{\text{th}} = 47.6 \text{ kV/cm}$  and  $J_{\text{th}} = 1.20 \text{ kA/cm}^2$ ). In Fig. 5a, we see that the



**Fig. 5** **a** Peak gain versus electric field. **b** Energy corresponding to peak gain as a function of electric field. Results are given for calculations using the full density-matrix (green circles), a semiclassical model (violet squares), and a comparison with NEGF (dashed blue line). Experimental results from Ref. [52] are denoted by a thick solid black line. The NEGF results are from Ref. [34] (Color figure online)

semiclassical model gives the same qualitative behavior as the density-matrix results. However, the gain is overestimated for all field values above threshold. The agreement between our DM results and the NEGF results is excellent, except for very high field values ( $> 70 \text{ kV/cm}$ ), where our DM model gives slightly higher gain.

Figure 5b shows the energy corresponding to the maximum gain in Fig. 5a. We see that both the DM and NEGF models give similar results, where the energetic position of the gain maximum is higher than the experimentally observed photon energy for all field values and the difference is greater at higher fields. However, both models agree with experiment slightly above threshold (52 kV/cm). The overestimation of the peak gain energy might look surprising considering the excellent agreement of the energy spacing between the upper and lower lasing level ( $\Delta_{8,7}$ ) shown in Fig. 2. However, other transitions such as 8 → 6 also contribute to gain (more pronounced at higher fields) and the energy spacing  $\Delta_{8,6}$  is overestimated by 3–13 meV. Another noticeable feature is the significant disagreement between the DM and semiclassical results at fields higher than 65 kV/cm. We attribute the difference to underestimation of the gain broadening using the semiclassical model, which leads to two distinct gain peaks, corresponding to  $\Delta_{8,7}$  and  $\Delta_{8,6}$ .

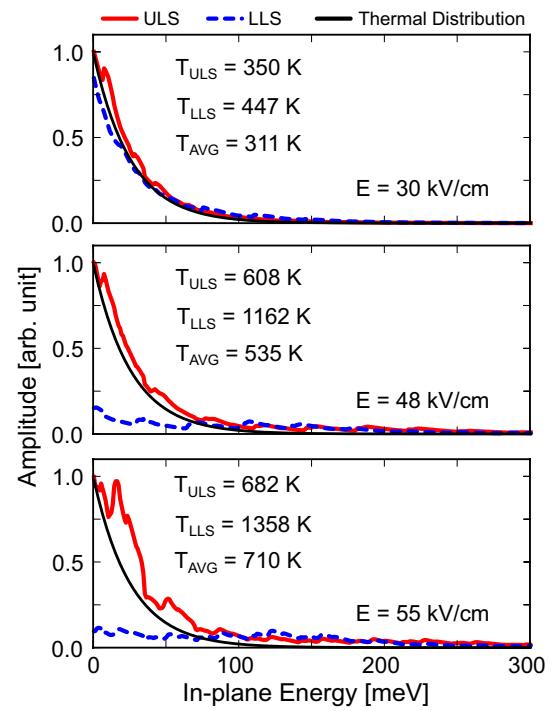
However, when using the full density matrix, the gain broadening is strong enough to combine the two peaks.

#### 4.1.4 In-plane dynamics

Figure 6 shows the in-plane distribution  $f_{N,0}^{E_k}$  for  $N$  corresponding to the upper lasing state (ULS) and lower lasing state (LLS). Results are shown far below threshold (30 kV/cm), at threshold (48 kV/cm), and above threshold (55 kV/cm). As a point of reference, a Maxwellian thermal distribution with a temperature of 300 K is also shown. For each value of electric field, the electron temperatures calculated using Eq. (18) are shown, as well as the weighted average for all subbands using (17). For low fields (30 kV/cm), the in-plane distribution is close to a Maxwellian distribution and the weighted average electron temperature is only 11 K higher than the lattice. For higher fields, significant electron heating takes place and in-plane distributions strongly deviate from a Maxwellian. There is a striking difference in shape and temperature between the ULS and LLS distributions at high fields. At threshold (48 kV/cm), the ULS is slightly (73 K) “hotter” than average, with a distribution that bears some resemblance to a thermal distribution, with most electrons having energy below 50 meV. However, the ULS in-plane distribution has a much thicker tail than the thermal distribution, resulting in a high electron temperature. The LLS is considerably hotter than average electron temperature, with a distribution that is flat up to around 150 meV (approximate lasing energy) and then slowly decays with a thick tail. These results are intuitive because the LLS has a low occupation and most of the occupation comes from electrons that recently scattered from the ULS with a large excess in-plane energy of  $\simeq 150$  meV (corresponding to about 1740 K), making the LLS very hot. On the other hand, electrons in the ULS come from long-lifetime states in the injector, where electrons have had time to emit many LO-phonons and get rid of excess energy.

We note that the inclusion of the optical-field feedback (not included in this work) is known to reduce the temperature of the LLS, because electrons can then transition from the ULS to the LLS via stimulated emission without acquiring any excess in-plane energy [53]. Inclusion of electron-electron scattering (also not included in this work) would also reduce the difference between the in-plane distributions of different subbands [54], even for the relatively low sheet doping density of  $n_{2D} = 7.8 \times 10^{10} \text{ cm}^{-2}$  in the considered device. However, we do not expect electron-electron interaction to significantly change the average electron temperature, because electron-electron scattering conserves the total energy of the electron system.

The electron temperatures shown in Fig. 6 are considerably higher than previously reported for the same device using NEGF [34]. For example, at a field of 55 kV/cm,



**Fig. 6** In-plane energy distribution for the ULS (solid red) and LLS (dashed blue) electrons. For comparison, a Maxwellian thermal distribution with a temperature of 300 K is also shown (thin black). Results are shown for electric fields of 30 kV/cm (top), 48 kV/cm (middle), and 55 kV/cm (bottom). For each value of the electric field, the electron temperatures of the ULS and LLS calculated using Eq. (18) are shown, as well as the weighted average electron temperature,  $T_{\text{avg}}$ , obtained using Eq. (17) (Color figure online)

we obtain  $T_{\text{ULS}} = 682$  K, compared with NEGF results of  $T_{\text{ULS}} = 398$  K [34]. The agreement is better for the LLS, where we obtain  $T_{\text{LLS}} = 1358$  K, while NEGF gives  $T_{\text{LLS}} = 1305$  K (private communications with the authors of Ref. [34]). Despite the disagreement, our results are similar to experimental and theoretical results for different QCL designs with similar wavelengths. For example, at threshold, electronic temperature of 800 K (at room temperature) has been experimentally estimated [55] for the GaAs-based QCL proposed in Ref. [56], which had a similar lasing wavelength  $\lambda \sim 9 \mu\text{m}$ . Even higher excess electron temperatures have been estimated by Monte Carlo calculations for a 9.4- $\mu\text{m}$  QCL, [54] where the authors obtained electron temperature of about 700 K at threshold for a lattice temperature of 77 K.

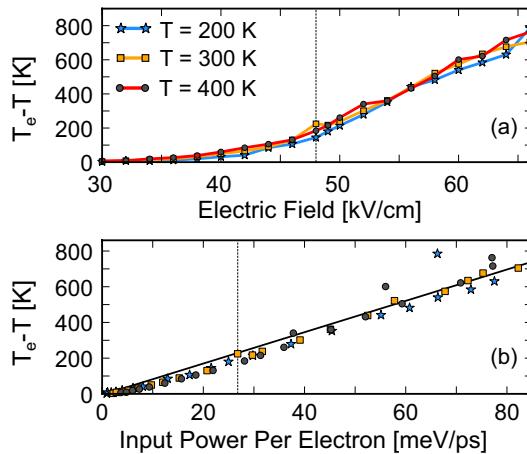
Figure 7a shows the average excess electron temperature  $T_e - T$  as a function of electric field for different values of lattice temperature  $T$ . The average electron temperature is calculated using Eq. (17). In Fig. 7a, we see that considerable electron heating takes place for fields higher than  $\sim 40$  kV/cm. For example, at a lattice temperature of 300 K, we observe an excess electron temperature of

225 K ( $T_e = 525$  K) around threshold (48.0 kV/cm) and a maximum value of 705 K ( $T_e = 1005$  K) for  $E = 66$  kV/cm.

Every time an electron traverses a single stage of length  $L_p$ , it gains energy  $\mathcal{E}_0$ , equal to the potential drop over a single period. It is therefore instructive to consider the excess electron temperature as a function of the normalized input power density

$$p_{\text{input}} = \frac{v}{L_p} \mathcal{E}_0 = \frac{JE}{n_{3D}} \quad (51)$$

where  $v$  is the drift velocity,  $E$  the electric field,  $J$  the current density and  $n_{3D}$  the average electron density. The normalized electrical power input gives the power transferred to each electron and is more convenient to compare between different devices than the (unnormalized) electrical power input  $P_{\text{input}} = JE$ , which depends linearly on the electron density (for low doping, as long as scattering rates and the Hartree potential are weakly affected). Figure 7b shows a scatter plot of the excess electron temperature as a function of the normalized input power for several values of the lattice temperature. The reason for a scatter plot is that the electron temperature is not a single-valued function of  $p_{\text{input}}$  because  $J$  is not an invertible function of in-plane energy ( $J$  is not a monotonically increasing function of electric field, as there are regions of negative differential resistance, so it cannot be inverted). Figure 7b shows that electron heating is well described with a simple energy-balance equation (EBE)



**Fig. 7** **a** Average excess electron temperature  $T_e - T$  as a function of field, where  $T_e$  is calculated using Eq. (18). Results are shown for  $T = 200$  K (blue stars),  $T = 300$  K (orange squares) and  $T = 400$  K (red circles). The dashed vertical line represent threshold field of 48 kV/cm for the room-temperature results. **b** Average excess electron temperature  $T_e - T$  as a function of electrical input power per electron, defined by Eq. (51). The black line shows a best fit with the energy-balance equation (52). The vertical line has the same meaning as in panel (a) (Color figure online)

$$T_e - T = \tau_E \frac{P_{\text{input}}}{k_B} \quad (52)$$

with an energy dissipation time  $\tau_E$ . A least-squares fit to Eq.(52) for the 300 K results gives  $\tau_E = 0.79$  ps, which is of the same order of magnitude as typical POP scattering times [57].

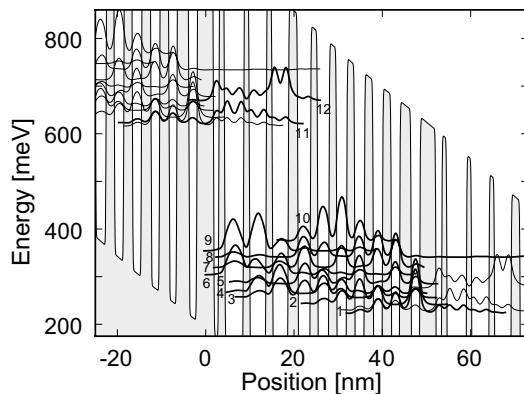
## 4.2 Strain-balanced design

In order to verify the accuracy of our model for QCLs based on strain-balanced heterostructures, we simulated the 4.6-μm QCL proposed in Ref. [58], with additional information in Ref. [59]. The device is based on alternating layers of  $\text{In}_{0.669}\text{Ga}_{0.331}\text{As}$  (well material) and  $\text{In}_{0.362}\text{Al}_{0.638}\text{As}$  (barrier material), with an approximate conduction-band discontinuity of 820 meV, enabling strong confinement of electrons at the considered wavelength of 4.6 μm (~270 meV). All material parameters relating to band structure (e.g., Kane energy) as well as other constants such as phonon energy and interface-roughness parameters are provided in “Appendix 6”.

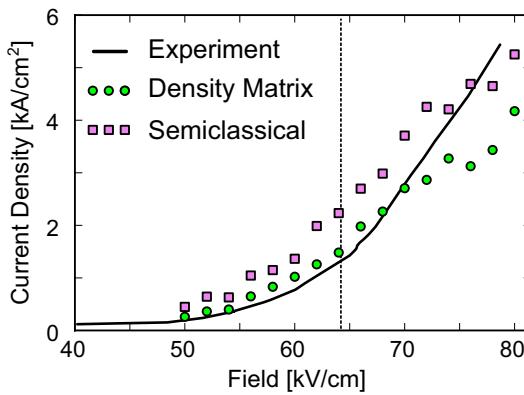
Figure 8 shows the conduction-band diagram (layer structure given in figure caption) and the probability density of the relevant eigenfunctions. The figure shows the 12 considered states belonging to a single period in bold and states belonging to neighboring periods are represented as thin curves. The lasing transition is between states 11 and 9, and state 12 is an unwanted leakage path. State number 1 is the energetically lowest state in the injector (often denoted as the ground state).

Figure 9 shows results for current density vs electric field calculated using the density-matrix model and a comparison with experiment as well as semiclassical results. We obtain excellent agreement between our density-matrix model and experiment up field strengths of 66 kV/cm, which is slightly above the experimentally determined threshold of 64 kV/cm. Above threshold, current density is underestimated due to the omission of the laser field in the theoretical results. Another possible reason for the underestimation of current density at high fields is lattice heating (experimental results are in continuous-wave operation), which is not included in this work. As in the case of the 8.5 μm QCL considered in Sect. 4.1, the semiclassical model overestimates current density for all field values. For example, around threshold (64 kV/cm), the density-matrix model predicts a current density of 1.48 kA/cm<sup>2</sup>, while the semiclassical model gives 2.23 kA/cm<sup>2</sup>, which is a relative difference of 40%. The biggest absolute difference between the density-matrix and semiclassical results is 1.57 kA/cm<sup>2</sup>, for a field strength of 76 kV/cm, while the biggest relative error is 56%, at a field strength of 52 kV/cm.

Figure 10 shows the peak gain as a function of electric field calculated using our density-matrix and semiclassical



**Fig. 8** Conduction-band edge (thin curve bounding gray area) and probability densities for the 12 eigenstates used in calculations (bold curves) for a field strength of 78 kV/cm. States belonging to neighboring periods are denoted by thin gray curves. The states are numbered in increasing order of energy, starting with the ground state in the injector. Using this convention, the lasing transition is from state 11 to state 9, and state 12 is an unwanted leakage path. The length of one period is 50.4 nm, with a layer structure (in nanometers), starting with the injector barrier (centered at the origin) **3.8/1.2 /1.3/4.3/1.3/3.8/1.4/3.6/2.2/2.8/1.7/2.5/1.8/2.2 /1.9/2.1/2.1/2.0/ 2.1/1.8/2.7/1.8**, with barriers denoted bold. Underlined layers are doped to  $1.1 \times 10^{17} \text{ cm}^{-3}$ , resulting in an average charge density of  $n_{3D} = 2.70 \times 10^{16} \text{ cm}^{-3}$  over a single stage



**Fig. 9** Calculated current density versus electric field obtained using our density-matrix (green circles) and semiclassical (violet squares) models. Also shown are experimental results (black curve) from Ref. [58]. Experimentally determined threshold field of 64 kV/cm is denoted by dashed vertical line. Both theoretical results are given for a lattice temperature of 300 K and experimental results are given for a heatsink temperature of 300 K. Experimental results are provided for continuous-wave operation, so some lattice heating is expected to take place for high current density and field (Color figure online)

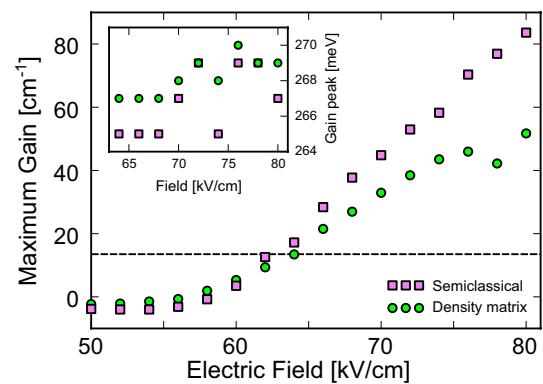
models. No estimate for the threshold gain  $G_{\text{th}}$  was given in Ref. [58] so the threshold gain (vertical dashed line) is simply the peak gain at the experimentally determined threshold field of 64 kV/cm, which gives a threshold gain of  $13.5 \text{ cm}^{-1}$ . From Fig. 10, we see that the density-matrix and semiclassical models give similar results up to threshold,

while for higher fields, the gain is overestimated by the semiclassical model. This is the same trend as for the 8.5- $\mu$ m-QCL considered in Sect. 4.1.

The inset of Fig. 10 shows the energy corresponding to the maximum gain for electric field strengths above threshold. From the inset we see that the photon energy corresponding to maximum gain depends weakly on field strength for both DM and semiclassical models, where energies range from 267 to 270 meV (4.60 to 4.65  $\mu$ m) using the DM model and 265 to 269 meV (4.62 to 4.69  $\mu$ m) using the semiclassical model, in very good agreement with experiment (4.6  $\mu$ m).

## 5 Conclusion

We derived a Markovian master equation for the single-electron density matrix that conserves the positivity of the density matrix and is applicable to electron transport in both mid-infrared and terahertz quantum cascade lasers. The master equation contains both the coherent (unitary) evolution of the density matrix as well as all relevant elastic and in-elastic scattering mechanism. Nonparabolicity in the band structure was accounted for by using a three-band  $\mathbf{k} \cdot \mathbf{p}$  model, which includes the conduction band, light-hole band, and the spin-orbit split-off band. The model is an improvement of previous work employing density matrices because it does not rely on a phenomenological treatment of dephasing across thick injection barriers [22] or a factorization of the density matrix into cross-plane and in-plane distribution. [60] We demonstrated the validity of the model by simulating QCLs based in



**Fig. 10** Peak gain vs electric field. Results are given for calculations both with off-diagonal density-matrix elements (green circles) and without them (violet squares). Dashed horizontal line shows the threshold gain  $G_{\text{th}} = 13.5 \text{ cm}^{-1}$ , calculated using the experimentally determined threshold field of 64 kV/cm. [58] The inset shows the energy position corresponding to peak gain as a function of electric field. All results are for a lattice temperature of 300 K (Color figure online)

lattice-matched as well as strain-balanced heterostructures, with lasing wavelengths of 8.5 μm and 4.6 μm, respectively. We compared our results to experiment as well as theoretical results based on NEGF and a semiclassical model.

For the 8.5-μm-QCL, we calculated the current density vs electric field strength and obtained excellent agreement with experiments for electric fields below threshold, while current density was underestimated above threshold due to our omission of the laser electromagnetic field. We obtained a good agreement with NEGF for low field values (< 42 kV/cm) and fair agreement for higher fields. At room temperature, we obtained a threshold electric field of  $E_{\text{th}} = 48.0 \text{ kV/cm}$  and threshold current density of  $J_{\text{th}} = 1.55 \text{ kA/cm}^2$ , in good agreement with experimental results of  $E_{\text{th}} = 48.0 \text{ kV/cm}$  and  $J_{\text{th}} = 1.50 \text{ kA/cm}^2$  and in fair agreement with NEGF results of  $E_{\text{th}} = 47.6 \text{ kV/cm}$  and  $J_{\text{th}} = 1.2 \text{ kA/cm}^2$ . Comparison with a semiclassical model shows that off-diagonal matrix elements play a significant role. The semiclassical model (off-diagonal matrix elements are ignored) was found to overestimate current density by 29% around threshold and up to 60% for other electric field values. Our results for maximum gain vs field were in good agreement with NEGF except for very high (> 70 kV/cm) field values. The semiclassical model was found to overestimate the peak gain above threshold, while giving fair agreement with DM and NEGF results below threshold. We predict significant electron heating, with in-plane distribution deviating strongly from a thermal distribution, especially the lower lasing state. We obtained an average electron temperature of 535 K around threshold, which is 235 K above the lattice temperature of 300 K. From a simple energy-balance equation, we obtained an energy-dissipation time of 0.79 ps for the considered device, which is similar to the scattering time due to polar-optical phonons.

For the 4.6-μm-QCL, we calculated current density vs electric field strength at room temperature and obtained excellent agreement with experiment for fields up to 66 kV/cm, which is slightly above the experimentally determined threshold of 64 kV/cm. Above threshold, the calculated current density was below experimental results due to our omission of the laser electric field as well effects of lattice heating in continuous-wave operation. As in the case of the 8.5-μm-QCL, We found that a semiclassical model overestimates current density for all field values. The semiclassical model was found to overestimate current density by 40% around threshold and up to 58% for other electric field values, demonstrating the need to include off-diagonal elements of the density matrix.

**Acknowledgements** This work was supported by the DOE-BES award DE-SC0008712 and by AFOSR award FA9550-18-1-0340. Preliminary

work was partially supported by the Splinter Professorship. Calculations were performed using the resources of the UW-Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences.

## Appendix 1: Electronic states

Within the envelope function approximation, the electron eigenfunctions near a band extremum  $\mathbf{k}_0$  ( $\Gamma$  valley in this work) can be written as a combination of envelope functions from different bands [61, 62]

$$\Psi_{n\mathbf{k}}(\mathbf{r}, z) = \sum_b \phi_{n,\mathbf{k}}^{(b)}(\mathbf{r}, z) u_{\mathbf{k}=\mathbf{k}_0}^{(b)}(\mathbf{r}, z), \quad (53)$$

where the  $b$  sum is over the considered bands,  $u_{\mathbf{k}_0}^{(b)}(\mathbf{k}, z)$  is a rapidly varying Bloch function that has the period of the lattice, and  $\phi_{n,\mathbf{k}}^{(b)}(\mathbf{r}, z)$  are slowly varying envelope functions belonging to band  $b$ . We make the approximation that the envelope functions factor into cross-plane and in-plane direction according to

$$\phi_{n,\mathbf{k}}^{(b)}(\mathbf{r}, z) = N_{\text{norm}} \psi_n^{(b)}(z) e^{i\mathbf{k}\cdot\mathbf{r}}, \quad (54)$$

where  $\psi_n^{(b)}(z) = \phi_{n,\mathbf{k}=0}^{(b)}(\mathbf{r}, z)$  is the envelope function for  $\mathbf{k} = 0$  and  $N_{\text{norm}}$  is a normalization constant. This separation of in- and cross-plane motion is a standard approximation when describing QCLs and other superlattices. [9] Without it, a separate eigenvalue problem for each  $\mathbf{k}$  would need to be calculated, which is very computationally expensive.

In order to model QCLs based on III-V semiconductors, accurate calculation the  $\Gamma$ -valley eigenstates is crucial. The simplest approach is the Ben Daniel–Duke model, in which only the conduction band is considered [61]. The Schrödinger equation for the conduction-band envelope function  $\psi_n^{(c)}(z)$  in Eq. (54) reduces to the 1D eigenvalue problem [9]

$$\left[ -\frac{\hbar^2}{2} \frac{d}{dz} \frac{1}{m^*(z)} \frac{d}{dz} + V^{(c)}(z) \right] \psi_n^{(c)}(z) = E_n \psi_n^{(c)}(z) \quad (55a)$$

$$E_{n,\mathbf{k}} = E_n + E_k, \quad (55b)$$

where  $V^{(c)}(z)$  is the conduction-band profile, and  $m^*(z)$  is the spatially dependent cross-plane effective mass of the conduction band. Typically, a parabolic dispersion relation  $E_k = \hbar^2 k^2 / (2m_{||}^*)$  is assumed. A single-band approximation such as Eq. (55b) can be justified when electrons have energy close to the conduction-band edge (compared with the bandgap  $E_{\text{gap}}$ ) and quantum well widths are larger than ~10 nm [61, 63]. This is the case for THz QCLs, where single-band approximations can give satisfactory results [42, 64]. However, in mid-IR QCLs, the well thicknesses can be

below 2 nm [52] and the energy of the lasing transition can exceed 150 meV, which is a significant fraction of typical bandgaps of the semiconductor materials, especially for low-bandgap materials containing InAs. In this case, including nonparabolicity in the electron Hamiltonian is important.

Nonparabolicity has previously been included in semiclassical [9, 53] density-matrix [34, 65], and NEGF-based [32, 34, 66] approaches using single-band models, where only the conduction band is modeled, but the influence of the valence bands is included via an energy-dependent effective mass. One of the most important effects of nonparabolicity is the energetic lowering of the high-energy states, such as the upper lasing level. This effect results from higher effective mass at high energies, reducing the kinetic energy term in the Schrödinger equation [67]. However, an energy-dependent effective mass has an undesirable trait in density-matrix and semiclassical models, the resulting eigenfunctions are not orthogonal [9]. In this work, we will calculate the band structure using a 3-band  $\mathbf{k} \cdot \mathbf{p}$  model, which includes the conduction (c), light hole (lh) and spin-orbit split-off (so) bands [68]. The three-band  $\mathbf{k} \cdot \mathbf{p}$  model has previously been used in semiclassical work on QCLs [16, 17]; present work is a quantum-mechanical QCL simulation relying this model.

The three-band Hamiltonian can be derived from an eight-band Hamiltonian for bulk zinc-blende crystals [69], which, at in-plane wave vector  $\mathbf{k} = 0$ , reduces to three-bands due to spin degeneracy and the heavy-hole band factoring out. The  $3 \times 3$  Hamiltonian for a strained bulk crystal can be written as a sum of three terms

$$H_B = H_{B,\text{edge}} + H_{B,\text{kin}} + H_{B,\text{strain}}. \quad (56)$$

The first term in Eq. (56) is diagonal and contains the band edges

$$H_{B,\text{edge}} = \begin{pmatrix} E_c & 0 & 0 \\ 0 & E_v & 0 \\ 0 & 0 & E_v - \Delta_{\text{so}} \end{pmatrix}, \quad (57)$$

where  $E_c$  ( $E_v$ ) is the conduction (valence) band edge and  $\Delta_{\text{so}}$  is the spin-orbit split-off energy.

The second term in Eq. (56) contains the kinetic contributions as well as coupling between the different bands which involve the cross-plane wavenumber  $k_z$

$$H_{B,\text{kin}} = \begin{pmatrix} (1+2F)\frac{\hbar^2 k_z^2}{2m_0} & \sqrt{\frac{\hbar^2 E_p}{3m_0}} k_z & \sqrt{\frac{\hbar^2 E_p}{6m_0}} k_z \\ \sqrt{\frac{\hbar^2 E_p}{3m_0}} k_z & -(\gamma_1 + 2\gamma_2)\frac{\hbar^2 k_z^2}{2m_0} & -\sqrt{2}\gamma_2 \frac{\hbar^2 k_z^2}{m_0} \\ \sqrt{\frac{\hbar^2 E_p}{6m_0}} k_z & -\sqrt{2}\gamma_2 \frac{\hbar^2 k_z^2}{m_0} & -\gamma_1 \frac{\hbar^2 k_z^2}{2m_0} \end{pmatrix}, \quad (58)$$

with  $E_p$  the Kane-energy, [70]  $m_0$  the free-electron mass,  $F$  is a correction to the conduction band effective mass due to higher energy  $\Gamma$  valleys [70],  $\gamma_1$  and  $\gamma_2$  are the modified Luttinger parameters [71] given by [69]

$$\begin{aligned} \gamma_1 &= \gamma_1^L - \frac{E_p}{3E_{\text{gap}} + \Delta_{\text{so}}} \\ \gamma_2 &= \gamma_2^L - \frac{1}{2} \frac{E_p}{3E_{\text{gap}} + \Delta_{\text{so}}}, \end{aligned} \quad (59)$$

where  $\gamma_1^L$  and  $\gamma_2^L$  are the standard Luttinger parameters [72] and  $E_{\text{gap}}$  is the  $\Gamma$ -valley bandgap. For a brief overview of the different material parameters, we refer the reader to Ref. [73].

The third term in Eq. (56) contains the effects of strain for a layer grown on a (001)-oriented (z-direction) substrate [62],

$$H_{B,\text{strain}} = \begin{pmatrix} A_\epsilon & 0 & 0 \\ 0 & -P_\epsilon + Q_\epsilon & \sqrt{2}Q_\epsilon \\ 0 & \sqrt{2}Q_\epsilon & -P_\epsilon \end{pmatrix}, \quad (60)$$

with

$$A_\epsilon = 2a_c \left( 1 - \frac{c_{12}}{c_{11}} \right) \frac{a_0 - a}{a} \quad (61a)$$

$$P_\epsilon = 2a_v \left( 1 - \frac{c_{12}}{c_{11}} \right) \frac{a_0 - a}{a} \quad (61b)$$

$$Q_\epsilon = -b \left( 1 + 2 \frac{c_{12}}{c_{11}} \right) \frac{a_0 - a}{a}, \quad (61c)$$

where  $a_c$ ,  $a_v$  and  $b$  are the Bir-Pikus deformation potentials [74] with the sign convention of all being negative,  $a_0$  and  $a$  are the lattice constant of the substrate and the grown material (in the absence of strain), respectively, and  $c_{11}$  and  $c_{12}$  are the elastic stiffness constants of the strained material. Numerical values for all material parameters used in this work are given in “Appendix 6”.

Equation (56) is valid for a bulk III-V semiconductors, where  $k_z$  refers to crystal momentum in the [001] direction. However, for a heterostructure under bias,  $k_z$  is no longer a good quantum number and must be replaced by a differential operator  $k_z \rightarrow -i \frac{\partial}{\partial z}$ . In addition, all the material properties become functions of position. Care must be taken when applying Eq. (56) to heterostructures because the proper operator symmetrization must be performed to preserve the hermiticity of the Hamiltonian. In this work, we follow the standard operator symmetrization procedure in  $\mathbf{k} \cdot \mathbf{p}$  theory for heterostructures [75]

$$\begin{aligned} f(z)k_z &\rightarrow \frac{1}{2}(f(z)k_z + k_z f(z)), \\ g(z)k_z^2 &\rightarrow k_z g(z)k_z, \end{aligned} \quad (62)$$

where  $f(z)$  and  $g(z)$  are arbitrary functions of position (e.g., Kane energy, or Luttinger parameter). The electron Hamiltonian for a heterostructure can then be written as

$$H_{\text{het}} = \begin{pmatrix} H_c & H_{c,lh} & H_{c,so} \\ H_{c,lh}^\dagger & H_{lh} & H_{lh,so} \\ H_{c,so}^\dagger & H_{lh,so}^\dagger & H_{so} \end{pmatrix}, \quad (63)$$

with

$$\begin{aligned} H_c &= E_c(z) - \frac{\hbar^2}{2m_0} \frac{\partial}{\partial z} (1 + 2F(z)) \frac{\partial}{\partial z} + A_\epsilon(z), \\ H_{lh} &= E_v(z) + \frac{\hbar^2}{2m_0} \frac{\partial}{\partial z} (\gamma_1(z) + 2\gamma_2(z)) \frac{\partial}{\partial z} \\ &\quad - P_\epsilon(z) + Q_\epsilon(z), \\ H_{so} &= E_v(z) - \Delta_{so}(z) + \frac{\hbar^2}{2m_0} \frac{\partial}{\partial z} \gamma_1(z) \frac{\partial}{\partial z} - P_\epsilon(z), \\ H_{c,lh} &= -\frac{i}{2} \sqrt{\frac{\hbar^2 E_p(z)}{3m_0}} \frac{\partial}{\partial z} - \frac{i}{2} \frac{\partial}{\partial z} \sqrt{\frac{\hbar^2 E_p(z)}{3m_0}}, \\ H_{c,so} &= -\frac{i}{2} \sqrt{\frac{\hbar^2 E_p(z)}{6m_0}} \frac{\partial}{\partial z} - \frac{i}{2} \frac{\partial}{\partial z} \sqrt{\frac{\hbar^2 E_p(z)}{6m_0}}, \\ H_{lh,so} &= \sqrt{2} \frac{\hbar^2}{m_0} \frac{\partial}{\partial z} \gamma_2(z) \frac{\partial}{\partial z} + \sqrt{2} Q_\epsilon(z). \end{aligned} \quad (64)$$

From Eq. (64), we see that the electron Hamiltonian now depends on spatial derivatives of material functions such as the Kane energy  $E_p(z)$  and modified Luttinger parameters  $\gamma_1(z)$ , and results will be sensitive to how an interface between two different materials is treated. One choice is to treat the material functions as piecewise-constant. However, with that choice, spatial derivatives become delta functions (and derivatives of delta functions) that are difficult to treat numerically. In this work, we will assume that material parameters are smooth functions that have slow variations on length scales shorter than a single monolayer  $L_{\text{ML}}$  (in heterostructures grown on InP,  $L_{\text{ML}} \approx 0.29$  nm at room temperature). The material parameters that we use are obtained using

$$f(z) = \frac{1}{\sqrt{\pi}\sigma} \int f_{\text{pc}}(z') e^{-(z-z')^2/\sigma^2} dz', \quad (65)$$

where we use  $\sigma = \frac{1}{2}L_{\text{ML}}$  and  $f_{\text{pc}}(z)$  is a piecewise constant material function that is discontinuous at material interfaces. As an example, Fig. 11 shows the Kane energy as well as the piecewise-constant variant for a single period of the QCL studied in Sect. 4.1, with  $\sigma = 0.29$  nm.

The eigenfunctions of the Hamiltonian  $H_{\text{het}}$  given in Eqs. (63)–(64) satisfy

$$H_{\text{het}} \overline{\psi_n}(z) = E_n \overline{\psi_n}(z), \quad (66)$$

where  $\overline{\psi_n}$  are three-dimensional vectors of envelope functions  $\overline{\psi_n}(z) = (\psi_n^c(z), \psi_n^{lh}(z), \psi_n^{so}(z))^T$ . We will use Dirac notation  $|n\rangle$ , with  $\langle z|n\rangle = \overline{\psi_n}(z)$  and  $|n, \mathbf{k}\rangle$  with  $\langle \mathbf{r}, z|n, \mathbf{k}\rangle = N_{\text{norm}} \overline{\psi_n}(z) e^{i\mathbf{k}\cdot\mathbf{r}}$  and normalization  $\langle n, \mathbf{k}|n', \mathbf{k}'\rangle = \delta_{n,n'} \delta(\mathbf{k} - \mathbf{k}')$ . The electron probability density is obtained by summing over the different bands

$$|\overline{\psi_n}(z)|^2 = |\langle z|n\rangle|^2 = \sum_b |\psi_n^b(z)|^2 \quad (67)$$

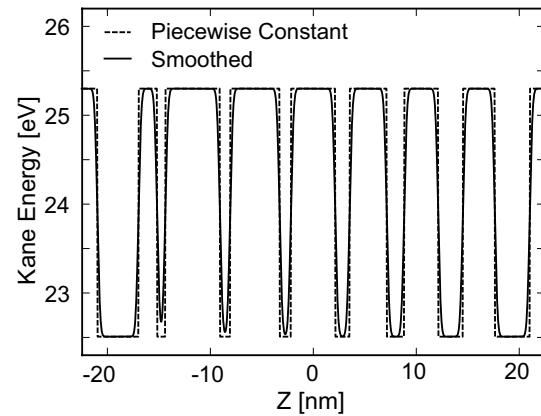
and matrix elements of an operator  $B$  such as  $\langle n|B|n'\rangle$  are understood as

$$\langle n|B|n'\rangle = \sum_b \int dz [\psi_n^{(b)}(z)]^* B^{(b)} \psi_{n'}^{(b)}(z), \quad (68)$$

where the sum is over the considered bands, with

$$B = \begin{pmatrix} B^{(c)} & 0 & 0 \\ 0 & B^{(lh)} & 0 \\ 0 & 0 & B^{(so)} \end{pmatrix}. \quad (69)$$

Note that  $B^{(b)}$  is an operation working on component  $b$  of the envelope function  $\overline{\psi_n}(z)$ . In this work, all operators are assumed to be identical for all three bands ( $B^{(c)} = B^{(lh)} = B^{(so)}$ ) except for the interaction potential due to interface roughness. The interaction potential due to interface roughness is not the same for all bands because the band



**Fig. 11** Kane energy as a function of position for a single period of the InGaAs/InAlAs-based mid-IR QCL described in Sect. 4.1. The dashed curve shows the piecewise constant version and the solid curve is the smoothed version given by Eq. (65), with  $\sigma = 0.29$  nm. This is double the value of  $\sigma$  we typically use and is chosen for visualization purposes. The region of higher Kane energy corresponds to InGaAs

discontinuities at material interfaces are not the same for different bands [see Appendix C.1].

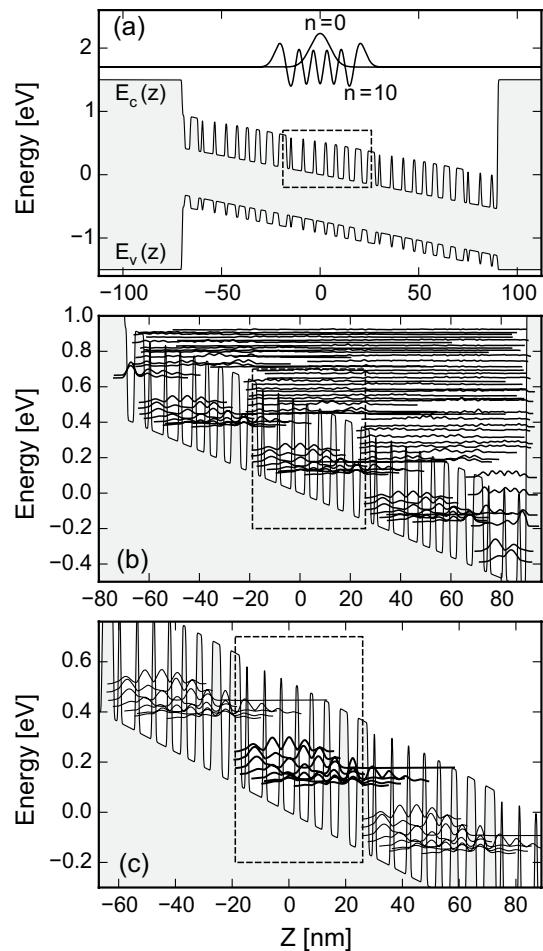
The eigenfunctions  $|n\rangle$  are calculated in a similar manner as in previous semiclassical work [9, 16, 17]. A typical computational domain is shown in Fig. 12(a), where the conduction band of a single period of the QCL from Ref. [52] is marked by a dashed rectangle. The computational domain contains 3–4 periods, padded with tall potential barriers, placed far away from the central region in order for the boundary conditions not to effect the eigenfunctions near the center of the computational domain. We solve the eigenvalue problem in (66) using a basis of Hermite functions (eigenfunctions of the harmonic oscillator), with two example basis functions shown in Fig. 12(a). We used Hermite functions because they are easy to compute and required less computational resources than finite-difference methods. We typically use a basis of  $\sim 400$  Hermite functions for each band, resulting in a matrix with dimension  $1200 \times 1200$ . The diagonalization procedure results in 1200 eigenstates, most of which are far above the conduction band or far below the valence bands. For transport calculations, we only include the states close to the conduction-band edge. This bound-state approach is known to produce spurious solutions, which are states with high amplitude far above the conduction band edge. In order to systematically single out the relevant states, we first “throw away” the states that have energies far ( $\sim 1$  eV) below or far above the conduction-band edge. This step typically reduces the number of states to  $\sim 100$ . Figure 12(b) shows the conduction-band edge along with the remaining eigenstates. Next, we remove the states that are obtained by translation of a state from the previous/next period (e.g., remove duplicates). We do accomplish this by calculating each eigenstate’s “center of mass”,

$$z_n = \int z |\psi_n(z)|^2 dz, \quad (70)$$

and only keep states within a range  $z \in [z_c, z_c + L_p]$ , where  $L_p$  is the period length. We will refer to this range as the center period. The choice of  $z_c$  is arbitrary and we typically use  $z_c = -L_p/4$ . Lastly, we follow the procedure in Ref. [9] and, for the remaining states, calculate the energy with respect to the conduction-band edge using

$$\tilde{E}_n = E_n - \int E_c(z) |\psi_n(z)|^2 dz. \quad (71)$$

States with  $\tilde{E}_n < 0$  are valence-band states and are discarded. We keep  $N_s$  states, with the lowest positive values of  $\tilde{E}_n$ . The value of  $N_s$  is chosen such that all eigenstates below the conduction band top are included. After this elimination step,  $N_s$  is the number of states used in transport calculations. The remaining states for  $N_s = 8$  are shown in bold in Fig. 12(c). States belonging to other periods (needed to



**Fig. 12** (a) Conduction band  $E_c(z)$  and valence band  $E_v(z)$  used for the calculation of eigenfunctions. Dashed rectangle shows a single period of the conduction band. Potential barriers are imposed to the far left and far right for both conduction and valence bands. Also show, are two of the Hermite functions ( $n = 0$  and  $n = 10$ ) used as basis functions. (b) After discarding the high- and low-energy states, about 80 states are left, most of which have an energy high above the conduction band or are greatly affected by the boundary conditions on the right. (c) After choosing 8 states with the lowest  $\tilde{E}_n$  and discarding copies from adjacent stages (see main text), the states in the central stage (bold curves) are left. States belonging to adjacent periods (thin curves) are calculated by translating states from the central stage in position and energy

calculate coupling between different periods) are finally obtained by translation in space and energy of the  $N_s$  states belonging to the center period. These states are identified by thin curves in Fig. 12(c).

## Appendix 2: Dissipator for electron–phonon interaction

Expanding the dissipator (commutators in Eq. (6)) gives 4 terms, which can be split into two Hermitian conjugate pairs

$$\begin{aligned} \mathcal{D}[\rho_e(t)] = & \int_0^\infty ds \text{Tr}_{\text{ph}} \{ \\ & + H_i \rho_e(t) \otimes \rho_{\text{ph}} e^{-i(H_0+H_{\text{ph}})s} H_i e^{i(H_0+H_{\text{ph}})s} \\ & - H_i e^{-i(H_0+H_{\text{ph}})s} H_i e^{i(H_0+H_{\text{ph}})s} \rho_e(t) \otimes \rho_{\text{ph}} \} \\ & + \text{h.c.} \end{aligned} \quad (72)$$

In order to proceed, we assume a Frölich-type Hamiltonian for the description of an electron interacting with a phonon bath in the single-electron approximation [76]

$$H_i = \frac{1}{\sqrt{V}} \sum_{\mathbf{Q}} \mathcal{M}(\mathbf{Q}) (b_{\mathbf{Q}} e^{i\mathbf{Q}\cdot\mathbf{R}} - b_{\mathbf{Q}}^\dagger e^{-i\mathbf{Q}\cdot\mathbf{R}}). \quad (73)$$

Here,  $b_{\mathbf{Q}}^\dagger$  ( $b_{\mathbf{Q}}$ ) is the creation (annihilation) operator for a phonon with wave vector  $\mathbf{Q}$  and the associated matrix element  $\mathcal{M}(\mathbf{Q})$  and  $V$  is the quantization volume. Note that the matrix elements are anti-Hermitian  $\mathcal{M}(\mathbf{Q})^* = -\mathcal{M}(\mathbf{Q})$ , making the interaction Hamiltonian Hermitian.

Inserting Eq. (73) into Eq. (72) gives

$$\begin{aligned} \mathcal{D} = & \frac{1}{V} \int_0^\infty ds \sum_{\mathbf{Q}, \mathbf{Q}'} \mathcal{M}(\mathbf{Q})^* \mathcal{M}(\mathbf{Q}') \text{Tr}_{\text{ph}} \{ \\ & - (b_{\mathbf{Q}} e^{i\mathbf{Q}\cdot\mathbf{R}} - b_{\mathbf{Q}}^\dagger e^{i\mathbf{Q}\cdot\mathbf{R}}) \rho_e(t) \otimes \rho_{\text{ph}} e^{-i(H_0+H_{\text{ph}})s} \times \\ & (b_{\mathbf{Q}'} e^{i\mathbf{Q}'\cdot\mathbf{R}} - b_{\mathbf{Q}'}^\dagger e^{-i\mathbf{Q}'\cdot\mathbf{R}}) e^{i(H_0+H_{\text{ph}})s} \\ & + (b_{\mathbf{Q}} e^{i\mathbf{Q}\cdot\mathbf{R}} - b_{\mathbf{Q}}^\dagger e^{-i\mathbf{Q}\cdot\mathbf{R}}) e^{-i(H_0+H_{\text{ph}})s} \times \\ & (b_{\mathbf{Q}'} e^{i\mathbf{Q}'\cdot\mathbf{R}} - b_{\mathbf{Q}'}^\dagger e^{-i\mathbf{Q}'\cdot\mathbf{R}}) e^{i(H_0+H_{\text{ph}})s} \rho_e(t) \otimes \rho_{\text{ph}} \} \\ & + \text{h.c..} \end{aligned} \quad (74)$$

In order to proceed, we use the fact that the thermal equilibrium-phonon density matrix is diagonal, so only terms with  $\mathbf{Q} = \mathbf{Q}'$  survive the trace over the phonon degree of freedom. We also use that only terms containing both creation and annihilation operators are nonzero after performing the trace. Using the two simplifications, we get

$$\begin{aligned} \mathcal{D} = & \frac{1}{V} \int_0^\infty ds \sum_{\mathbf{Q}} |\mathcal{M}(\mathbf{Q})|^2 \text{Tr}_{\text{ph}} \{ \\ & + b_{\mathbf{Q}} e^{i\mathbf{Q}\cdot\mathbf{R}} \rho_e(t) \otimes \rho_{\text{ph}} e^{-iH_0s} e^{-iH_{\text{ph}}s} b_{\mathbf{Q}}^\dagger e^{iH_{\text{ph}}s} e^{-i\mathbf{Q}\cdot\mathbf{R}} e^{iH_0s} \\ & + b_{\mathbf{Q}}^\dagger e^{-i\mathbf{Q}\cdot\mathbf{R}} \rho_e(t) \otimes \rho_{\text{ph}} e^{-iH_0s} e^{-iH_{\text{ph}}s} b_{\mathbf{Q}} e^{iH_{\text{ph}}s} e^{i\mathbf{Q}\cdot\mathbf{R}} e^{iH_0s} \\ & - b_{\mathbf{Q}} e^{i\mathbf{Q}\cdot\mathbf{R}} e^{-iH_0s} e^{-iH_{\text{ph}}s} b_{\mathbf{Q}}^\dagger e^{iH_{\text{ph}}s} e^{-i\mathbf{Q}\cdot\mathbf{R}} e^{iH_0s} \rho_e(t) \otimes \rho_{\text{ph}} \\ & - b_{\mathbf{Q}}^\dagger e^{-i\mathbf{Q}\cdot\mathbf{R}} e^{-iH_0s} e^{-iH_{\text{ph}}s} b_{\mathbf{Q}} e^{iH_{\text{ph}}s} e^{i\mathbf{Q}\cdot\mathbf{R}} e^{iH_0s} \rho_e(t) \otimes \rho_{\text{ph}} \} \\ & + \text{h.c..} \end{aligned} \quad (75)$$

Performing the trace, and going from a sum over  $\mathbf{Q}$  to an integral, we get

$$\begin{aligned} \mathcal{D} = & \frac{1}{(2\pi)^3} \sum_{\pm} \int_0^\infty ds \int d^3 Q |\mathcal{M}(\mathbf{Q})|^2 \\ & \times \left\{ \left( \frac{1}{2} \mp \frac{1}{2} + N_{\mathbf{Q}} \right) e^{\mp iE_0 s} e^{i\mathbf{Q}\cdot\mathbf{R}} \rho_e(t) e^{-iH_0 s} e^{-i\mathbf{Q}\cdot\mathbf{R}} e^{iH_0 s} \right. \\ & \left. - \left( \frac{1}{2} \mp \frac{1}{2} + N_{\mathbf{Q}} \right) e^{\pm iE_0 s} e^{-i\mathbf{Q}\cdot\mathbf{R}} e^{-iH_0 s} e^{i\mathbf{Q}\cdot\mathbf{R}} e^{iH_0 s} \rho_e(t) \right\} \\ & + \text{h.c.}, \end{aligned} \quad (76)$$

where  $E_{\mathbf{Q}}$  refers to the energy of a phonon with wave vector  $\mathbf{Q}$  and we have defined the expectation value of the phonon number operator  $\text{Tr}_{\text{ph}} \{ \rho_{\text{ph}} b_{\mathbf{Q}}^\dagger b_{\mathbf{Q}} \} = N_{\mathbf{Q}}$ . Note that the upper (lower) sign refers to absorption (emission) terms. In the following, we will assume phonons are dispersionless, so  $E_{\mathbf{Q}} = E_0$  and  $N_{\mathbf{Q}} = N_0 = (e^{E_0/k_B T} - 1)^{-1}$ ,  $T$  is the lattice temperature, and  $k_B$  is the Boltzmann constant. In order to keep expressions more compact, we define the *scattering weight*

$$\mathcal{W}^\pm(\mathbf{Q}) = \left( \frac{1}{2} \mp \frac{1}{2} + N_0 \right) |\mathcal{M}(\mathbf{Q})|^2. \quad (77)$$

We can then write the dissipator due to a single dispersionless phonon type as (after reintroduction of  $\hbar$ ),

$$\begin{aligned} \mathcal{D} = & \frac{1}{\hbar^2 (2\pi)^3} \sum_{\pm} \int_0^\infty ds \int d^3 Q \{ \\ & + \mathcal{W}^\pm(\mathbf{Q}) e^{\mp iE_0 s/\hbar} e^{i\mathbf{Q}\cdot\mathbf{R}} \rho_e(t) e^{-iH_0 s/\hbar} e^{-i\mathbf{Q}\cdot\mathbf{R}} e^{iH_0 s/\hbar} \\ & - \mathcal{W}^\pm(\mathbf{Q}) e^{\pm iE_0 s/\hbar} e^{-i\mathbf{Q}\cdot\mathbf{R}} e^{-iH_0 s/\hbar} e^{i\mathbf{Q}\cdot\mathbf{R}} e^{iH_0 s/\hbar} \rho_e(t) \} \\ & + \text{h.c.}, \end{aligned} \quad (78)$$

where the total dissipator will be obtained later by summing over the different relevant phonon branches. The dissipator can be written as  $\mathcal{D} = \mathcal{D}^{\text{in}} + \mathcal{D}^{\text{out}}$ , corresponding to the positive and negative terms in (78), respectively. We will derive the form of the matrix elements of the in-scattering term. The out-scattering term is less complicated and can be obtained by inspection of the in-scattering term and only the final results will be given.

By using the completeness relation four times, we can write

$$\begin{aligned} \mathcal{D}^{\text{in}} = & \frac{1}{\hbar^2 (2\pi)^3} \int d^3 Q \int_0^\infty ds \sum_{n_{1234}, \pm} \int d^2 k_{1234} \\ & \times \langle n_1, \mathbf{k}_1 | e^{i\mathbf{Q}\cdot\mathbf{R}} | n_2, \mathbf{k}_2 \rangle e^{\mp iE_0 s/\hbar} \mathcal{W}^\pm(\mathbf{Q}) \\ & \times \langle n_3, \mathbf{k}_3 | e^{-iH_0 s/\hbar} e^{-i\mathbf{Q}\cdot\mathbf{R}} e^{iH_0 s/\hbar} | n_4, \mathbf{k}_4 \rangle \\ & \times \langle n_2, \mathbf{k}_2 | \rho_e | n_3, \mathbf{k}_3 \rangle | n_1, \mathbf{k}_1 \rangle \langle n_4, \mathbf{k}_4 | + \text{h.c.}, \end{aligned} \quad (79)$$

where  $\sum n_{1234}$  and  $\int d^2 k_{1234}$  refer to summation over  $n_1$  through  $n_4$  and integration of  $\mathbf{k}_1$  through  $\mathbf{k}_4$ , respectively. Integration of  $\mathbf{k}_2$  through  $\mathbf{k}_4$  and shifting the integration variable  $\mathbf{Q} \rightarrow -\mathbf{Q} + \mathbf{k}_1$  gives

$$\begin{aligned} \mathcal{D}^{\text{in}} = & \frac{1}{\hbar^2(2\pi)^3} \int d^3Q \int_0^\infty ds \sum_{n_{1234}, \pm} \int d^2k_1 \\ & \times \mathcal{W}^\pm(\mathbf{Q} - \mathbf{k}_1) (n_1|n_2)_{Q_z} (n_3|n_4)_{Q_z}^* \rho_{n_2, n_3}^{E_q} \\ & \times e^{-i\frac{s}{\hbar}(A_{n_3, n_4} - E_{k_1} \pm E_0 + E_q)} |n_1, \mathbf{k}_1\rangle \langle n_4, \mathbf{k}_1| + \text{h.c.}, \end{aligned} \quad (80)$$

where  $E_k$  is the in-plane dispersion relation,  $q = \sqrt{Q_x^2 + Q_y^2}$  is the magnitude of the in-plane component of  $\mathbf{Q}$ ,  $A_{nm} = E_n - E_m$  is the energy spacing of the cross-plane eigenfunctions and

$$(n|m)_{Q_z} = \sum_b \int_{-\infty}^\infty [\psi_n^{(b)}(z)]^* e^{iQ_z z} \psi_m^{(b)}(z) dz, \quad (81)$$

where the sum is over the considered bands (conduction, light-hole and split-off bands in this work). In order to perform the  $s$  integration, we use

$$\int_0^\infty e^{-i\Delta s/\hbar} = \pi \hbar \delta(\Delta) - i\hbar \mathcal{P} \frac{1}{\Delta}, \quad (82)$$

where  $\mathcal{P}$  denotes the Cauchy principal value, which causes a small shift to energies called the *Lamb shift*. [44] By omitting the Lamb shift term and dropping the subscript on the  $\mathbf{k}_1$  integration variable, we get

$$\begin{aligned} \mathcal{D}^{\text{in}} = & \frac{\pi}{\hbar(2\pi)^3} \sum_{n_{1234}, \pm} \int d^2k \int d^3Q \mathcal{W}^\pm(\mathbf{Q} - \mathbf{k}) \\ & (n_1|n_2)_{Q_z} (n_3|n_4)_{Q_z}^* \rho_{n_2, n_3}^{E_k + A_{n_4, n_3} \mp E_0} \\ & \delta[E_q - (E_k + A_{n_4, n_3} \mp E_0)] |n_1, \mathbf{k}\rangle \langle n_4, \mathbf{k}| + \text{h.c..} \end{aligned} \quad (83)$$

Multiplying by  $\langle N|$  from the left and  $|N + M\rangle$  from the right and renaming the sum variables  $n_2 \rightarrow n$  and  $n_3 \rightarrow m$  gives the matrix elements of the in-scattering contribution of the dissipator, giving

$$\begin{aligned} \left( \frac{\partial f_{N,M}^{E_k}}{\partial t} \right)^{\text{in}} = & \frac{\pi}{\hbar(2\pi)^3} \sum_{n,m,\pm} \int d^3Q \mathcal{W}^\pm(\mathbf{Q} - \mathbf{k}) \\ & (N|n)_{Q_z} (m|N + M)_{Q_z}^* \times \rho_{n,m}^{E_k + A_{N+M,m} \mp E_0} \\ & \delta[E_q - (E_k + A_{N+M,m} \mp E_0)] + \text{h.c..} \end{aligned} \quad (84)$$

The sums involving  $n$  and  $m$  are over all integers. However, it is convenient due to numerical reasons to shift the sum variables in such a way that terms with small  $|n|$  and  $|m|$  dominate, and terms with large  $|n|$  and  $|m|$  approach zero. This can be done by making the switch  $n \rightarrow N + n$  and  $m \rightarrow N + M + m$ , giving

$$\begin{aligned} \left( \frac{\partial f_{N,M}^{E_k}}{\partial t} \right)^{\text{in}} = & \frac{\pi}{\hbar(2\pi)^3} \sum_{n,m,\pm} \int d^3Q \mathcal{W}^\pm(\mathbf{Q} - \mathbf{k}) \times \\ & (N|N + n)_{Q_z} (N + M + m|M + M)_{Q_z}^* \times \\ & \rho_{N+n, N+M+m}^{E_k + A_{N+M,N+M+m} \mp E_0} \delta[E_q - (E_k + A_{N+M,N+M+m} \mp E_0)] \\ & + \text{h.c..} \end{aligned} \quad (85)$$

In order to write the RHS in Eq. (85) in terms of  $f_{N,M}^{E_k}$ , we use

$$\rho_{N+n, N+M+m}^{E_k} = \rho_{N+n, N+n+(M+m-n)}^{E_k} = f_{N+n, M+m-n}^{E_k} \quad (86)$$

and get

$$\begin{aligned} \left( \frac{\partial f_{N,M}^{E_k}}{\partial t} \right)^{\text{in}} = & \frac{\pi}{\hbar(2\pi)^3} \sum_{n,m,\pm} \int d^3Q \mathcal{W}^\pm(\mathbf{Q} - \mathbf{k}) \\ & (N|N + n)_{Q_z} (N + M + m|M + M)_{Q_z}^* \times \\ & f_{N+n, M+m-n}^{E_k + A_{N+M,N+M+m} \mp E_0} \delta[E_q - (E_k + A_{N+M,N+M+m} \mp E_0)] \\ & + \text{h.c..} \end{aligned} \quad (87)$$

or

$$\begin{aligned} \left( \frac{\partial f_{N,M}^{E_k}}{\partial t} \right)^{\text{in}} = & \sum_{n,m,\pm} \Gamma_{NMnmE_k}^{\pm, \text{in}} f_{N+n, M+m-n}^{E_k + A_{N+M,N+M+m} \mp E_0} \\ & + \text{h.c.}, \end{aligned} \quad (88)$$

where we have defined the in-scattering rates

$$\begin{aligned} \Gamma_{NMnmE_k}^{\pm, \text{in}} = & \frac{\pi}{\hbar(2\pi)^3} \int d^3Q \mathcal{W}^\pm(\mathbf{Q} - \mathbf{k}) \\ & (N|N + n)_{Q_z} (N + M + m|M + M)_{Q_z}^* \times \\ & \delta[E_q - (E_k + A_{N+M,N+M+m} \mp E_0)]. \end{aligned} \quad (89)$$

Note that the RHS in Eq. (89) looks like it depends on the direction of  $\mathbf{k}$ . However, the scattering weight only depends on the magnitude  $|\mathbf{Q} - \mathbf{k}|$ , so the coordinate system for the  $\mathbf{Q}$  integration can always been chosen relative to  $\mathbf{k}$ , which means the RHS only depends on the magnitude  $|\mathbf{k}|$ .

In order to proceed, we need to specify the in-plane dispersion relation  $E_k$ , as well as the scattering mechanism, which determines the form of the scattering weight  $\mathcal{W}^\pm(\mathbf{Q} - \mathbf{k})$ . In this work, we will assume a parabolic dispersion relation  $E_k = \hbar^2 k^2 / 2m_{||}^*$ , with a single in-plane effective mass  $m_{||}^*$ , defined as the average effective mass of the  $N_s$  considered eigenstates

$$\frac{1}{m_{||}^*} = \frac{1}{N_s} \sum_{n=1}^{N_s} \sum_b \int \frac{|\psi_n^{(b)}(z)|^2}{m_b^*(z)} dz, \quad (90)$$

where  $m_b^*(z)$  is the position-dependent effective mass of band  $b$ , calculated using the procedure given in “Appendix 6”. The approximation of a parabolic in-plane dispersion relation can fail in short-wavelength QCLs, where electrons can have in-plane energies comparable to the material bandgap [53]. It is possible to go beyond a parabolic dispersion relation by including a term proportional to  $k^4$  (in-plane nonparabolicity). However, the inclusion of a  $k^4$  term would significantly complicate the integration over  $q$  in Eq. (89) and is beyond the scope of this work.

Since we do not have an analytical expression for the eigenfunctions, the products  $(n|m)_{Q_z}$  have to be calculated numerically. For a more compact notation, we define

$$I_{n,m,k,\ell}(Q_z) = (n|m)_{Q_z} (k|\ell)_{Q_z}^*. \quad (91)$$

Switching to cylindrical coordinates  $d^3Q \rightarrow qdqd\theta dQ_z$ , making the change of variables  $E_q = \hbar^2 q^2 / 2m_\parallel^*$  and integrating over  $E_q$  gives

$$\begin{aligned} \Gamma_{NMnmE_k}^{\pm,\text{in}} &= \frac{m_\parallel^*}{2\pi\hbar^3} u[E_k + \Delta_{N+M,N+M+m} \mp E_0] \\ &\times \int_0^\infty dQ_z I_{N,N+n,N+M+m,N+M}(Q_z) \\ &\times \left[ \frac{1}{2\pi} \int_0^{2\pi} d\theta \mathcal{W}^\pm(\mathbf{Q} - \mathbf{k}) \right]_{E_q=E_k+\Delta_{N+M,N+M+m}\mp E_0}, \end{aligned} \quad (92)$$

where the term in the square brackets is the scattering weight averaged over the angle between  $\mathbf{Q}$  and  $\mathbf{k}$ , with the in-plane part of  $\mathbf{Q}$  constrained by energy conservation, and  $u[x]$  is the Heaviside function, which is zero when no positive value of  $E_q$  can satisfy energy conservation. Note that we have also taken advantage of the fact that the real part of the integrand is even in  $Q_z$ , allowing us to limit the range of integration to positive  $Q_z$ . The resulting expressions from the angle-averaged scattering weight can be quite cumbersome, so we define the angle-averaged scattering weight

$$\tilde{\mathcal{W}}^\pm(E_k, Q_z, E_q) = \frac{1}{2\pi} \int_0^{2\pi} d\theta \mathcal{W}^\pm(\mathbf{Q} - \mathbf{k}), \quad (93)$$

simplifying the in-scattering rate to

$$\begin{aligned} \Gamma_{NMnmE_k}^{\pm,\text{in}} &= \frac{m_\parallel^*}{2\pi\hbar^3} u[E_k + \Delta_{N+M,N+M+m} \mp E_0] \\ &\times \int_0^\infty dQ_z I_{N,N+n,N+M+m,N+M}(Q_z) \\ &\times \tilde{\mathcal{W}}^\pm(E_k, Q_z, E_k + \Delta_{N+M,N+M+m} \mp E_0). \end{aligned} \quad (94)$$

In order to proceed, we need to specify the scattering weight, which depends on the kind of phonon interaction being considered. Calculations of scattering rates for longitudinal

acoustic phonons and polar optical phonons are performed in “Appendix 4”.

### Appendix 3: Dissipator for elastic scattering mechanisms

In addition to electron–phonon scattering, there are numerous elastic scattering mechanisms that are relevant to QCLs. The most important ones are interface roughness, ionized impurities, and alloy scattering [9, 31, 33]. The derivation of the corresponding scattering rates is slightly different from the case for phonons and will be demonstrated for the case of interface roughness, which is a very important elastic scattering mechanism for both THz and mid-IR QCLs [50, 77, 78] and has been found to decrease the upper lasing state lifetime by a factor of 2 at room temperature in mid-IR QCLs [78]. For other elastic scattering mechanisms, we only briefly discuss the interaction potential and give final expressions for rates in “Appendix 4”.

#### Interface roughness

For an electron in band  $b$ , deviation from a perfect interface between two different semiconductor material can be modeled using an interaction Hamiltonian on the form [79]

$$H_{\text{IR}}^{(b)} = \sum_i \Delta_i(\mathbf{r}) \frac{\partial V^{(b)}}{\partial z} \Big|_{z=z_i} \simeq \sum_i \Delta V_i^{(b)} \delta(z - z_i) \Delta_i(\mathbf{r}), \quad (95)$$

where  $i$  labels an interface located at position  $z_i$ ,  $V^{(b)}(z)$  the band edge corresponding to band  $b$  and  $\Delta V_i^{(b)}$  is the band discontinuity at interface  $i$ . The function  $\Delta_i(\mathbf{r})$  represents the deviation from a perfect interface at different in-plane positions  $\mathbf{r}$ . For the conduction band of an InGaAs/InAlAs heterostructure, the sign of  $\Delta V_i^{(b)}$  is positive when going from well material to barrier material (from the left) and negative when going from barrier to well material. For the light-hole and split-off bands, the signs are opposite (with respect to the conduction band).

The introduction of the interaction Hamiltonian in Eq. (95) breaks translational invariance in the  $xy$ -plane, and the resulting density matrix will no longer be diagonal in the in-plane wave vector. In order to get around this limitation, we follow the treatment in Refs. [9, 31, 80] and only consider statistical properties of the interface roughness, which is contained in the correlation function

$$\langle \Delta_i(\mathbf{r}) \Delta_j(\mathbf{r}') \rangle = \delta_{ij} C(|\mathbf{r} - \mathbf{r}'|) \quad (96)$$

where  $\langle \dots \rangle$  denotes an average over the macroscopic interface area,  $C(|\mathbf{r}|)$  is the spatial autocorrelation function and the Kronecker delta function means that we neglect correlations between different interfaces. In this work, we employ a

Gaussian correlation function  $C(|\mathbf{r}|) = \Delta^2 e^{-|\mathbf{r}|^2/\Lambda^2}$ , characterized by standard deviation  $\Delta$  and correlation length  $\Lambda$ . However, calculations will be taken as far as possible without specifying a correlation function. When calculating rates, we will mainly use its Fourier transform,

$$C(|\mathbf{q}|) = \int d^2 C(|\mathbf{r}|) e^{-i\mathbf{r} \cdot \mathbf{q}}. \quad (97)$$

The derivation for the interface-roughness scattering rates proceeds the same way as for electron–phonon scattering until Eq. (72). For the case of interface roughness, there is no phonon degree of freedom to trace over and we get

$$\begin{aligned} \mathcal{D}_{\text{IR}} &= \int_0^\infty ds \{ +H_{\text{IR}} \rho_e(t) e^{-iH_0 s} H_{\text{IR}} e^{iH_0 s} \\ &\quad - H_{\text{IR}} e^{-iH_0 s} H_{\text{IR}} e^{iH_0 s} \rho_e(t) \} + \text{h.c.} \end{aligned} \quad (98)$$

After using the completeness relation 4 times and performing the  $s$  integration, the in-scattering term in Eq. (98) becomes

$$\begin{aligned} \mathcal{D}_{\text{IR}}^{\text{in}} &= \frac{\pi}{\hbar} \sum_{n_{1234}} \int d^2 k_{1234} \delta[E(n_4, \mathbf{k}_4) - E(n_3, \mathbf{k}_3)] \\ &\quad \langle n_2, \mathbf{k}_2 | \rho_e | n_3, \mathbf{k}_3 \rangle \times \langle n_1, \mathbf{k}_1 | H_{\text{IR}} | n_2, \mathbf{k}_2 \rangle \\ &\quad \langle n_3, \mathbf{k}_3 | H_{\text{IR}} | n_4, \mathbf{k}_4 \rangle | n_1, \mathbf{k}_1 \rangle \langle n_4, \mathbf{k}_4 | + \text{h.c..} \end{aligned} \quad (99)$$

Integrating over  $\mathbf{k}_3$  gives

$$\begin{aligned} \mathcal{D}_{\text{IR}}^{\text{in}} &= \frac{\pi}{\hbar} \sum_{n_{1234}} \int d^2 k_{124} \delta[E(n_4, \mathbf{k}_4) - E(n_3, \mathbf{k}_2)] \rho_{n_2, n_3}^{E_{k_2}} \\ &\quad \times \langle n_1, \mathbf{k}_1 | H_{\text{IR}} | n_2, \mathbf{k}_2 \rangle \langle n_3, \mathbf{k}_2 | H_{\text{IR}} | n_4, \mathbf{k}_4 \rangle \\ &\quad \times | n_1, \mathbf{k}_1 \rangle \langle n_4, \mathbf{k}_4 | + \text{h.c.} \end{aligned} \quad (100)$$

Now let us focus on the part of the integrand containing the interaction Hamiltonian

$$\begin{aligned} &\langle n_1, \mathbf{k}_1 | H_{\text{IR}} | n_2, \mathbf{k}_2 \rangle \langle n_3, \mathbf{k}_2 | H_{\text{IR}} | n_4, \mathbf{k}_4 \rangle \\ &= \sum_{ij} (n_1 | n_2 )_{z_i, \text{IR}} (n_3 | n_4 )_{z_j, \text{IR}}^* \\ &\quad \times \int d^2 r d^2 r' \langle \mathbf{r} | \mathbf{k}_1 \rangle^* \langle \mathbf{r} | \mathbf{k}_2 \rangle \langle \mathbf{r}' | \mathbf{k}_2 \rangle^* \langle \mathbf{r}' | \mathbf{k}_4 \rangle \Delta_i(\mathbf{r}) \Delta_j(\mathbf{r}'), \end{aligned} \quad (101)$$

where we have defined

$$(n|m)_{z_i, \text{IR}} = \sum_b \Delta V_i^{(b)} \psi_n^*(z_i) \psi_m(z_i), \quad (102)$$

where the sum is over the conduction (c), light hole (lh), and spin-orbit split-off bands (so). We use a convention where the band discontinuity  $\Delta V_i^{(b)}$  is positive for the conduction band. In order to proceed, we replace the product  $\Delta_i(\mathbf{r}) \Delta_j(\mathbf{r}')$  with the Fourier decomposition of its spatial average using Eqs. (96)–(97) and get

$$\begin{aligned} &\langle n_1, \mathbf{k}_1 | H_{\text{IR}} | n_2, \mathbf{k}_2 \rangle \langle n_3, \mathbf{k}_2 | H_{\text{IR}} | n_4, \mathbf{k}_4 \rangle \\ &= \sum_i (n_1 | n_2 )_{z_i, \text{IR}} (n_3 | n_4 )_{z_i, \text{IR}}^* \frac{1}{(2\pi)^2} \\ &\quad \int d^2 q \delta[\mathbf{k}_2 - (\mathbf{k}_1 - \mathbf{q})] \delta[\mathbf{k}_4 - (\mathbf{k}_2 + \mathbf{q})] C(|\mathbf{q}|). \end{aligned} \quad (103)$$

Inserting Eq. (103) into Eq. (100) and integrating over  $\mathbf{k}_4$  and  $\mathbf{k}_2$ , dropping the subscript index of  $\mathbf{k}_1$  and shifting the integration variable  $\mathbf{q} \rightarrow -\mathbf{q} + \mathbf{k}$  gives

$$\begin{aligned} \mathcal{D}_{\text{IR}}^{\text{in}} &= \frac{\pi}{\hbar(2\pi)^2} \sum_{n_{1234}} \int d^2 k \int d^2 q \delta[E_q - \Delta_{n_4, n_3} - E_k] \\ &\quad \sum_i (n_1 | n_2 )_{z_i, \text{IR}} (n_3 | n_4 )_{z_i, \text{IR}}^* C(|\mathbf{q} - \mathbf{k}|) \\ &\quad \rho_{n_2, n_3}^{E_q} | n_1, \mathbf{k} \rangle \langle n_4, \mathbf{k}_1 | + \text{h.c..} \end{aligned} \quad (104)$$

Switching to polar coordinates  $d^2 q \rightarrow q dq d\theta$  and performing the change of variables  $E_q = \hbar^2 q^2 / (2m_{||}^*)$  gives

$$\begin{aligned} \mathcal{D}_{\text{IR}}^{\text{in}} &= \frac{m_{||}^*}{2\hbar^3} \sum_{n_{1234}} \int d^2 k \sum_i (n_1 | n_2 )_{z_i, \text{IR}} (n_3 | n_4 )_{z_i, \text{IR}}^* \\ &\quad \times \frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty dE_q \delta[E_q - \Delta_{n_4, n_3} - E_k] \\ &\quad \times C(|\mathbf{q} - \mathbf{k}|) \rho_{n_2, n_3}^{E_q} | n_1, \mathbf{k} \rangle \langle n_4, \mathbf{k}_1 | + \text{h.c..} \end{aligned} \quad (105)$$

Performing the  $E_q$  integration, renaming the dummy variables  $n_3 \rightarrow m$  and  $n_2 \rightarrow n$  and multiplying from the left by  $\langle N |$  and from the right with  $|N + M\rangle$  gives

$$\begin{aligned} \left( \frac{\partial f_{N,M}^{E_k}}{\partial t} \right)_{\text{IR}}^{\text{in}} &= \frac{m_{||}^*}{2\hbar^3} \sum_{n,m} u [\Delta_{N+M,m} + E_k] \\ &\quad \times \sum_i (N | n )_{z_i, \text{IR}} (m | N + M )_{z_i, \text{IR}}^* \\ &\quad \times \tilde{C}(E_k, \Delta_{N+M,m} + E_k) \rho_{nm}^{\Delta_{N+M,m} + E_k} + \text{h.c..} \end{aligned} \quad (106)$$

where we have defined the angle-averaged correlation function

$$\tilde{C}(E_k, E_q)$$

$$= \frac{1}{2\pi} \int_0^{2\pi} d\theta C \left( \sqrt{\frac{2m_{||}^*}{\hbar^2}} \sqrt{E_k + E_q - 2\sqrt{E_k E_q} \cos(\theta)} \right). \quad (107)$$

Shifting the sum variables  $n \rightarrow N + n$ ,  $m \rightarrow N + M + m$  and using Eq. (86), we can write

$$\begin{aligned} \left( \frac{\partial f_{N,M}^{E_k}}{\partial t} \right)_{\text{IR}}^{\text{in}} &= \frac{m_{\parallel}^*}{2\hbar^3} \sum_{n,m} u[\Delta_{N+M,N+M+m} + E_k] \\ &\times \sum_i (N|N+n)_{z_i,\text{IR}} (N+M+m|N+M)_{z_i,\text{IR}}^* \\ &\times \tilde{C}(E_k, \Delta_{N+M,N+M+m} + E_k) f_{N+n,M+m-n}^{\Delta_{N+M,N+M+m} + E_k} + \text{h.c.} \\ &= 2 \sum_{n,m} I_{NMnmE_k}^{\text{in},\text{IR},\pm} f_{N+n,M+m-n}^{\Delta_{N+M,N+M+m} + E_k} + \text{h.c.}, \end{aligned} \quad (108)$$

where we have defined  $I_{NMnmE_k}^{\text{in},\text{IR},\pm}$ , the in-scattering rate due to interface roughness. The purpose of the factor of two in the definition of the rate is so we can write

$$\left( \frac{\partial f_{N,M}^{E_k}}{\partial t} \right)_{\text{IR}}^{\text{in}} = \sum_{n,m,\pm} I_{NMnmE_k}^{\text{in},\text{IR},\pm} f_{N+n,M+m-n}^{\Delta_{N+M,N+M+m} + E_k} + \text{h.c..} \quad (109)$$

In other words, we have split the interface roughness scattering into two identical absorption and emission terms to be consistent with the notation for phonon scattering in “Appendix 2”. In order to calculate the rates, we need to choose a correlation function. The rates are calculated in “Appendix 4” for a Gaussian correlation function, which is often used to model QCLs [9, 34]. The reason we choose a Gaussian correlation function is that it results in a simple analytical expression for the angle-averaged correlation function  $\tilde{C}$ , given by Eq. (107). Another popular choice is an exponential correlation function [9, 31, 33]. However, angle-averaged expressions for the correlation function are more complicated, involving elliptical integrals [9, 31].

## Ionized impurities

The effects of the long-range Coulomb interaction from ionized impurities are already included to lowest order with a mean-field treatment (Poisson’s equation). The goal of this section is to include higher order effects by including scattering of electrons with ionized impurities. We assume a 3D doping density of the form

$$N(z) = N_{2D} \delta(z - z_{\ell}), \quad (110)$$

which is a delta-doped sheet at cross-plane position  $z_{\ell}$ . The advantage of Eq. (110) is that an arbitrary doping density can be approximated by a linear combination of closely spaced delta-doped sheets. Typically, in QCLs, only one [64] or a few [52, 59] layers are doped and the doping is kept away from the active region to minimize nonradiative transitions from the upper to lower lasing level. Note that, in this context, a layer refers to a barrier or a well, not atomic layers. The corresponding scattering rates due to the ionized-impurity scattering distribution in Eq. (110) is calculated in Appendix E.4.

## Random alloy scattering

In ternary semiconductor alloys  $A_xB_{1-x}C$ , the energy of band  $b$  can be calculated using the virtual crystal approximation, [79] where the alloy band energy is obtained by averaging over the band profiles of the two materials  $AC$  and  $BC$

$$V_{b,\text{VCA}}(\mathbf{R}) = \sum_n [xV_{b,\text{AC}}(\mathbf{R} - \mathbf{R}_n) + (1-x)V_{b,\text{BC}}(\mathbf{R} - \mathbf{R}_n)], \quad (111)$$

where the sum is over all unit cells  $n$  and  $V_{b,\text{AC}}(\mathbf{R} - \mathbf{R}_n)$  and  $V_{b,\text{BC}}(\mathbf{R} - \mathbf{R}_n)$  are the corresponding bands of the AC and BC material, respectively. The difference between an alloy and a binary compound is that, in alloys, periodicity of the crystal is broken by the random distribution of  $AC$  and  $BC$  unit cells. This deviation from a perfect lattice can be modeled as a scattering potential, which can be considered much weaker than Eq. (111). The scattering potential is given by the difference between the real band energy and the band energy obtained using the virtual crystal approximation. We will follow the treatment of alloy scattering in Refs. [9, 33, 81] and write the scattering potential (which we assume is the same for all bands) as

$$V_{\text{alloy}}(\mathbf{R}) = \sum_n C_n^x \delta(\mathbf{R} - \mathbf{R}_n), \quad (112)$$

where the sum is over all unit cells and  $C_n^x$  is a random variable, analogous to the interface-roughness correlation function in Eq. (96), with

$$\begin{aligned} \langle C_n^x \rangle &= 0 \\ \langle C_n^x C_m^x \rangle &= \delta_{nm} x_n (1 - x_n) (V_n \Omega)^2, \end{aligned} \quad (113)$$

where  $\Omega = a^3/4$  is equal to one-fourth of the volume of a unit cell and  $V_n$  the alloy scattering potential in unit cell  $n$ , which can be approximated as the conduction band offset of the two binary materials, corresponding to  $x = 0$  and  $x = 1$ . The presence of the Kronecker delta function  $\delta_{nm}$  means that the location of A and B atoms are assumed to be completely uncorrelated. The derivation of corresponding scattering rates proceeds in the same way as for interface roughness scattering and results are given in “Appendix 4”. The final expressions for the rates are considerably simpler than the ones due to interface roughness due to the local nature of the interaction potential in Eq. (112).

## Appendix 4: Is the dissipator really of the Lindblad form?

For each of the scattering mechanisms considered in this work, it can be shown that the corresponding dissipator is of the Lindblad form [43]. In fact, each dissipator was derived according to a procedure that guarantees a Lindbladian dissipator [42, 44, 82, 83]. However, one can readily convince oneself of this fact. In what follows, we prove the Lindblad form for the electron–phonon interaction dissipator as an example.

According to Eq. (78), we can write the dissipator due to a single dispersionless phonon type as

$$\begin{aligned} \mathcal{D} = & \frac{1}{\hbar^2(2\pi)^3} \sum_{\pm} \int_0^\infty ds \int d^3Q \{ \\ & + \mathcal{W}^\pm(\mathbf{Q}) e^{\mp iE_0 s/\hbar} e^{i\mathbf{Q}\cdot\mathbf{R}} \rho_e(t) e^{-iH_0 s/\hbar} e^{-i\mathbf{Q}\cdot\mathbf{R}} e^{iH_0 s/\hbar} \\ & - \mathcal{W}^\pm(\mathbf{Q}) e^{\pm iE_0 s/\hbar} e^{-i\mathbf{Q}\cdot\mathbf{R}} e^{-iH_0 s/\hbar} e^{i\mathbf{Q}\cdot\mathbf{R}} e^{iH_0 s/\hbar} \rho_e(t) \} \\ & + \text{h.c.}, \end{aligned} \quad (114)$$

The dissipator can be written as a summation of in- and out-scattering terms ( $\mathcal{D} = \mathcal{D}^{\text{in}} + \mathcal{D}^{\text{out}}$ ), corresponding to the positive and negative terms in Eq. (78), respectively. By using the completeness relation four times, we can write

$$\begin{aligned} \mathcal{D}^{\text{in}} = & \frac{1}{\hbar^2(2\pi)^3} \int d^3Q \int_0^\infty ds \sum_{n_{1234}, \pm} \int d^2k_{1234} \\ & \times \langle n_1, \mathbf{k}_1 | e^{i\mathbf{Q}\cdot\mathbf{R}} | n_2, \mathbf{k}_2 \rangle e^{\mp iE_0 s/\hbar} \mathcal{W}^\pm(\mathbf{Q}) \\ & \times \langle n_2, \mathbf{k}_2 | \rho_e | n_3, \mathbf{k}_3 \rangle \\ & \times \langle n_3, \mathbf{k}_3 | e^{-iH_0 s/\hbar} e^{-i\mathbf{Q}\cdot\mathbf{R}} e^{iH_0 s/\hbar} | n_4, \mathbf{k}_4 \rangle | n_1, \mathbf{k}_1 \rangle \langle n_4, \mathbf{k}_4 | + \text{h.c.}, \end{aligned} \quad (115)$$

where  $\sum n_{1234}$  and  $\int d^2k_{1234}$  refer to summation over  $n_1$  through  $n_4$  and integration of  $\mathbf{k}_1$  through  $\mathbf{k}_4$ , respectively. Since,  $\rho$  and  $\mathcal{D}$  are diagonal in  $\mathbf{k}$ , this can be further simplified to

$$\begin{aligned} \mathcal{D}^{\text{in}} = & \frac{1}{\hbar^2(2\pi)^3} \int d^3Q \int_0^\infty ds \sum_{n_{1234}, \pm} \int d^2k_{1,2} \\ & \times \langle n_1, \mathbf{k}_1 | e^{i\mathbf{Q}\cdot\mathbf{R}} | n_2, \mathbf{k}_2 \rangle e^{-i(\Delta_{n_3,n_4} \pm E_0 + E_{\mathbf{k}_2} - E_{\mathbf{k}_1})s/\hbar} \mathcal{W}^\pm(\mathbf{Q}) \\ & \times \langle n_2, \mathbf{k}_2 | \rho_e | n_3, \mathbf{k}_3 \rangle \\ & \times \langle n_3, \mathbf{k}_3 | e^{-i\mathbf{Q}\cdot\mathbf{R}} | n_4, \mathbf{k}_4 \rangle | n_1, \mathbf{k}_1 \rangle \langle n_4, \mathbf{k}_4 | + \text{h.c.}. \end{aligned} \quad (116)$$

By defining

$$\mathcal{A}_Q = \langle n_1, \mathbf{k}_1 | e^{+i\mathbf{Q}\cdot\mathbf{R}} | n_2, \mathbf{k}_2 \rangle | n_1, \mathbf{k}_1 \rangle \langle n_2, \mathbf{k}_2 |, \quad (117)$$

$$\Gamma_{n_1 n_2}^{\mathbf{k}_1 \mathbf{k}_2} = \frac{\mathcal{W}^\pm(\mathbf{Q})}{\hbar^2(2\pi)^3} \int_0^\infty ds e^{-i(\Delta_{n_1,n_2} \pm E_0 + E_{\mathbf{k}_2} - E_{\mathbf{k}_1})s/\hbar}, \quad (118)$$

Equation (116) can be written compactly as

$$\mathcal{D}^{\text{in}} = 2 \sum_{n_{1,2,3,4}} \int d^3Q \int d^2\mathbf{k}_{1,2} \Gamma \mathcal{A} \rho \mathcal{A}^\dagger. \quad (119)$$

Note that  $\Gamma$ , as defined in Eq. (118), is Hermitian (transposition of the top two or bottom two indices results in the complex conjugate of the original). Similarly, the out-scattering terms can be written as

$$\begin{aligned} \mathcal{D}^{\text{out}} = & -\frac{1}{\hbar^2(2\pi)^3} \int d^3Q \int_0^\infty ds \sum_{n_{1234}, \pm} \int d^2k_{1234} \\ & \times \langle n_1, \mathbf{k}_1 | e^{-i\mathbf{Q}\cdot\mathbf{R}} | n_2, \mathbf{k}_2 \rangle e^{i(\Delta_{n_4,n_3} \pm E_0 + E_{\mathbf{k}_1} - E_{\mathbf{k}_2})s/\hbar} \mathcal{W}^\pm(\mathbf{Q}) \\ & \times \langle n_2, \mathbf{k}_2 | e^{+i\mathbf{Q}\cdot\mathbf{R}} | n_3, \mathbf{k}_1 \rangle \\ & \times \langle n_3, \mathbf{k}_2 | \rho_e | n_4, \mathbf{k}_2 \rangle | n_1, \mathbf{k}_1 \rangle \langle n_4, \mathbf{k}_1 | + \text{h.c.}. \end{aligned} \quad (120)$$

Using Eqs. (117) and (118),  $\mathcal{D}^{\text{out}}$  is written as

$$\mathcal{D}^{\text{out}} = - \sum_{n_{1,2,3,4}} \int d^3Q \int d^2\mathbf{k}_{1,2} \Gamma (\mathcal{A}^\dagger \mathcal{A} \rho + \rho \mathcal{A}^\dagger \mathcal{A}),. \quad (121)$$

Summing in- and out-scattering terms would give us the total dissipator for electron–phonon interaction of the form

$$\mathcal{D} = \sum_{n_{1,2,3,4}} \int d^3Q \int d^2\mathbf{k}_{1,2} \Gamma (2\mathcal{A} \rho \mathcal{A}^\dagger - \{\mathcal{A}^\dagger \mathcal{A}, \rho\}), \quad (122)$$

which is of the Lindblad form.

## Appendix 5: Calculation of scattering rates

### Longitudinal acoustic phonons

We will start with longitudinal acoustic (LA) phonons, which have the following scattering weight for absorption [79]

$$\mathcal{W}_{\text{LA}}^+(\mathbf{Q}) = N_Q \frac{\hbar^2 D_{\text{ac}}^2 Q^2}{2\rho E_Q} = (e^{E_Q/kT} - 1)^{-1} \frac{D_{\text{ac}}^2 E_Q}{2\rho v_s^2}, \quad (123)$$

where we have used the LA phonon dispersion relation  $E_Q = \hbar v_s Q$ , where  $v_s$  is the speed of sound in the medium,  $\rho$  the mass density and  $D_{\text{ac}}$  is the acoustic deformation potential. We cannot directly use the expression in Eq. (123) because the phonon energy is not constant, which was an approximation we used to derive Eq. (94). It is possible to discretize the relevant range of  $Q$  and treat LA phonons with different magnitude  $Q$  as independent scattering mechanism. However, we would need to calculate and store the corresponding rates for each value of  $Q$ , which is prohibitive in terms of CPU time as well as memory use. Another option is to consider the limit  $E_Q \ll kT$  and treat the LA phonons as an elastic scattering mechanism. However, because the

LA phonons can exchange an arbitrarily small energy with the lattice, they play the important role of low energy thermalization at energies much smaller than the polar optical phonon. A second reason to include a nonzero energy for the LA phonons is that they tend to smooth the in-plane energy distribution of the density matrix, greatly facilitating numerical convergence. For this reason, we follow the treatment in Ref. [31] and approximate acoustic phonons as having a single energy  $E_{\text{LA}} \ll kT$ . Typically, we will assume the LA phonon energy is equal to the energy spacing in the simulation (0.5–2 meV). Using this approximation, we get

$$\mathcal{W}_{\text{LA}}^{\pm}(\mathbf{Q}) = \left( \frac{1}{2} \mp \frac{1}{2} + N_{\text{LA}} \right) \frac{D_{\text{ac}}^2 E_{\text{LA}}}{2\rho v_s^2}, \quad (124)$$

with  $N_{\text{LA}} = (\exp(E_{\text{LA}}/kT) - 1)^{-1}$ . The expression in Eq. (124) is isotropic (does not depend on  $\mathbf{Q}$ ), so the angular integration in Eq. (94) is trivial and we get

$$\begin{aligned} \Gamma_{NMnmE_k}^{\pm, \text{in,LA}} &= \left( \frac{1}{2} \mp \frac{1}{2} + N_{\text{LA}} \right) \frac{D_{\text{ac}}^2 E_{\text{LA}} m_{\parallel}^*}{4\pi\hbar^3 \rho v_s^2} \\ &\times u[E_k + \Delta_{N+M,N+M+m} \mp E_{\text{LA}}] \\ &\int_0^\infty dQ_z I_{N,N+n,N+M+m,N+M}(Q_z). \end{aligned} \quad (125)$$

The corresponding out scattering rates are

$$\begin{aligned} \Gamma_{NMnmE_k}^{\pm, \text{out,LA}} &= \left( \frac{1}{2} \mp \frac{1}{2} + N_{\text{LA}} \right) \frac{D_{\text{ac}}^2 E_{\text{LA}} m_{\parallel}^*}{4\pi\hbar^3 \rho v_s^2} \\ &\times u[E_k + \Delta_{N+n,N+M+m} \pm E_{\text{LA}}] \\ &\times \int_0^\infty dQ_z I_{N+M,N+M+m,N+M+m,N+n}(Q_z). \end{aligned} \quad (126)$$

## Polar longitudinal-optical phonons

The scattering weight for polar longitudinal-optical (LO) phonons is [79]

$$\mathcal{W}_{\text{LO}}^{\pm}(\mathbf{Q}) = \left( \frac{1}{2} \mp \frac{1}{2} + N_{\text{LO}} \right) \beta_{\text{LO}} \frac{Q^2}{(Q^2 + Q_{\text{D}}^2)^2}, \quad (127)$$

with

$$\beta_{\text{LO}} = \frac{e^2 E_{\text{LO}}}{2\varepsilon_0} \left( \frac{1}{\varepsilon_s^\infty} - \frac{1}{\varepsilon_s^0} \right), \quad (128)$$

where  $\varepsilon_0$  is the permittivity of vacuum,  $\varepsilon_s^\infty$  and  $\varepsilon_s^0$  are the high-frequency and low-frequency relative bulk permittivities, respectively,  $E_{\text{LO}}$  is the polar-optical phonon energy and  $Q_{\text{D}}$  is the Debye screening wave vector defined by  $Q_{\text{D}}^2 = n_{3D} e^2 / (\varepsilon_s^\infty k_B T)$ , with  $n_{3D}$  the average 3D electron density in the device. We note that QCLs are typically

low-doped, so the effect of screening is minimal. However, the screening facilitates numerical calculations by avoiding the singularity when  $|\mathbf{Q}| = 0$  in Eq. (127). Integration over the angle between  $\mathbf{Q}$  and  $\mathbf{k}$  gives

$$\begin{aligned} \widetilde{\mathcal{W}}_{\text{LO}}^{\pm}(E_k, Q_z, E_q) &= \frac{1}{2\pi} \int_0^{2\pi} \frac{|\mathbf{Q} - \mathbf{k}|^2}{(|\mathbf{Q} - \mathbf{k}|^2 + Q_{\text{D}}^2)^2} \\ &= \frac{1}{2\pi} \int_0^{2\pi} d\theta \frac{Q_z^2 + q^2 + k^2 - 2qk \cos(\theta)}{(Q_z^2 + q^2 + k^2 - 2qk \cos(\theta) + Q_{\text{D}}^2)^2} \\ &= \frac{1}{Q_z^2 + q^2 + k^2} \\ &\times \left[ 1 + \frac{Q_{\text{D}}^2}{Q_z^2 + q^2 + k^2} - \frac{4k^2 q^2}{(Q_z^2 + q^2 + k^2)^2} \right] \\ &\times \left[ \left( 1 + \frac{Q_{\text{D}}^2}{Q_z^2 + q^2 + k^2} \right)^2 - \frac{4k^2 q^2}{(Q_z^2 + q^2 + k^2)^2} \right]^{-\frac{3}{2}}, \end{aligned} \quad (129)$$

which can be written as

$$\begin{aligned} \widetilde{\mathcal{W}}_{\text{LO}}^{\pm}(E_k, Q_z, E_q) &= \frac{\hbar^2}{2m^*} \frac{\left( \frac{1}{2} \mp \frac{1}{2} + N_{\text{LO}} \right) \beta_{\text{LO}}}{E_z + E_k + E_q} \\ &\times \left[ 1 + \frac{E_{\text{D}}}{E_z + E_q + E_k} - \frac{4E_k E_q}{(E_z + E_q + E_k)^2} \right] \\ &\times \left[ \left( 1 + \frac{E_{\text{D}}}{E_z + E_q + E_k} \right)^2 - \frac{4E_k E_q}{(E_z + E_q + E_k)^2} \right]^{-\frac{3}{2}}, \end{aligned} \quad (130)$$

with  $E_z = \hbar^2 Q_z^2 / (2m_{\parallel}^*)$ . Note that this definition of  $E_z$  is only for convenience, it has nothing to do with energy quantization in the  $z$ -direction. The in- and out-scattering rates can now be written as

$$\begin{aligned} \Gamma_{NMnmE_k}^{\pm, \text{in,LO}} &= \frac{m_{\parallel}^*}{2\pi\hbar^3} u[E_k + \Delta_{N+M,N+M+m} \mp E_{\text{LO}}] \\ &\times \int_0^\infty dQ_z I_{N,N+n,N+M+m,N+M}(Q_z) \\ &\times \widetilde{\mathcal{W}}_{\text{LO}}^{\pm}(E_k, Q_z, E_k + \Delta_{N+M,N+M+m} \mp E_{\text{LO}}) \end{aligned} \quad (131)$$

$$\begin{aligned} \Gamma_{NMnmE_k}^{\pm, \text{out,LO}} &= \frac{m_{\parallel}^*}{2\pi\hbar^3} u[E_k + \Delta_{N+n,N+M+m} \pm E_{\text{LO}}] \\ &\times \int_0^\infty dQ_z I_{N+M,N+M+m,N+M+m,N+n}(Q_z) \\ &\times \widetilde{\mathcal{W}}_{\text{LO}}^{\pm}(E_k, Q_z, E_k + \Delta_{N+n,N+M+m} \pm E_{\text{LO}}), \end{aligned} \quad (132)$$

with the angle-averaged scattering weights  $\widetilde{\mathcal{W}}_{\text{LO}}^{\pm}$  given by Eq. (130).

## Interface roughness

In order to calculate the scattering rates due to interface roughness in Eq. (108), we need to calculate the angle-averaged correlation function  $\tilde{C}(E_k, E_q)$ . In this work, we employ a Gaussian correlation function  $C(|\mathbf{r}|) = \Delta^2 e^{-|\mathbf{r}|^2/\Lambda^2}$ , with a corresponding Fourier transform  $C(|\mathbf{q}|) = \pi \Delta^2 \Lambda^2 e^{-|\mathbf{q}|^2 \Lambda^2/4}$ . The angle-averaged correlation function can be calculated using

$$\begin{aligned} \tilde{C}(E_k, E_q) &= \frac{1}{2\pi} \int_0^{2\pi} C(|\mathbf{q} - \mathbf{k}|) d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} d\theta C\left(\sqrt{\frac{m_{||}^*}{2\hbar^2}} \sqrt{E_k + E_q - 2\sqrt{E_k E_q} \cos(\theta)}\right) \\ &= \Delta^2 \Lambda^2 \int_0^\pi d\theta e^{-\frac{m_{||}^* \Lambda^2}{2\hbar^2} (E_k + E_q - 2\sqrt{E_k E_q} \cos(\theta))} \\ &= \pi \Delta^2 \Lambda^2 e^{-\frac{m_{||}^* \Lambda^2}{2\hbar^2} (E_k + E_q)} I_0(\sqrt{E_k E_q} m_{||}^* \Lambda^2 / \hbar^2), \end{aligned} \quad (133)$$

where we have used Eq. (3.339) from Ref. [84]

$$\int_0^\pi e^{-x \cos(\theta)} d\theta = \pi I_0(x), \quad (134)$$

with  $I_0(x)$ , the modified Bessel function of the first kind of order zero. The in-scattering rates can now be written as

$$\begin{aligned} \Gamma_{NMnmE_k}^{\text{in,IR},\pm} &= \frac{m_{||}^*}{4\hbar^3} u[\Delta_{N+M,N+M+m} + E_k] \\ &\times \sum_i (N|N+n)_{z_i,\text{IR}} (N+M+m|N+M)_{z_i,\text{IR}}^* \\ &\times \tilde{C}(E_k, \Delta_{N+M,N+M+m} + E_k), \end{aligned}$$

with the corresponding out-scattering rates

$$\begin{aligned} \Gamma_{NMnmE_k}^{\text{out,IR},\pm} &= \frac{m_{||}^*}{4\hbar^3} u[\Delta_{N+n,N+M+m} + E_k] \\ &\times \sum_i (N+M|N+M+m)_{z_i,\text{IR}} (N+M+m|N+n)_{z_i,\text{IR}}^* \\ &\times \tilde{C}(E_k, \Delta_{N+n,N+M+m} + E_k), \end{aligned} \quad (135)$$

where we have split the rates into identical absorption (+) and emission (-) terms.

## Ionized impurities

The potential due to a single ionized impurity situated at  $\mathbf{R}_i$ , can be written as

$$V(\mathbf{R}) = -\frac{e^2}{4\pi\epsilon} \frac{e^{-\beta|\mathbf{R}-\mathbf{R}_i|}}{|\mathbf{R}-\mathbf{R}_i|}, \quad (136)$$

where  $\beta$  is the inverse screening length. Now suppose there are many impurities, situated on a sheet  $\mathbf{R}_i = (\mathbf{r}_i, z_\ell)$ . The resulting interaction Hamiltonian can be written as

$$\begin{aligned} H_{\text{II}} &= -\frac{e^2}{4\pi\epsilon} \sum_i \frac{e^{-\beta\sqrt{(\mathbf{r}-\mathbf{r}_i)^2+(z-z_\ell)^2}}}{\sqrt{(\mathbf{r}-\mathbf{r}_i)^2+(z-z_\ell)^2}} \\ &= \frac{-e^2}{8\pi^2\epsilon} \sum_i \int d^2 q e^{i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}_i)} \frac{e^{-\sqrt{\beta^2+q^2}|z-z_\ell|}}{\sqrt{\beta^2+q^2}}, \end{aligned} \quad (137)$$

where the second line in Eq. (137) is the 2D Fourier decomposition of the first line. We now proceed as we did with interface roughness scattering up until Eq. (100) and get the in-scattering part of the dissipator due to ionized-impurities

$$\begin{aligned} \mathcal{D}_{\text{II}}^{\text{in}} &= \frac{\pi}{\hbar} \sum_{n_{1234}} \int d^2 k_{124} \delta[E(n_4, \mathbf{k}_4) - E(n_3, \mathbf{k}_2)] \rho_{n_2, n_3}^{E_{k_2}} \\ &\times \langle n_1, \mathbf{k}_1 | H_{\text{II}} | n_2, \mathbf{k}_2 \rangle \langle n_3, \mathbf{k}_2 | H_{\text{II}} | n_4, \mathbf{k}_4 \rangle | n_1, \mathbf{k}_1 \rangle \langle n_4, \mathbf{k}_4 | \\ &+ \text{h.c..} \end{aligned} \quad (138)$$

Inserting Eq. 137 into Eq. (138), we can simplify the integrand using

$$\begin{aligned} &\langle n_1, \mathbf{k}_1 | H_{\text{II}} | n_2, \mathbf{k}_2 \rangle \langle n_3, \mathbf{k}_2 | H_{\text{II}} | n_4, \mathbf{k}_4 \rangle \\ &= \frac{e^4}{64\pi^4\epsilon^2} \sum_{r_i, r_j} \int d^2 q d^2 q' \frac{1}{\sqrt{\beta^2+q^2}} \frac{1}{\sqrt{\beta^2+q'^2}} \\ &\times e^{-i\mathbf{q}\cdot\mathbf{r}_i} e^{-i\mathbf{q}'\cdot\mathbf{r}_j} (n_1|n_2)_{q,\ell}^{\text{II}} (n_2|n_3)_{q',\ell}^{\text{II}} \\ &\times \langle \mathbf{k}_1 | \mathbf{k}_2 + \mathbf{q} \rangle \langle \mathbf{k}_2 | \mathbf{k}_4 - \mathbf{q}' \rangle. \end{aligned} \quad (139)$$

with

$$(n|m)_{q,\ell}^{\text{II}} = \sum_b \int dz [\psi_n^{(b)}(z)]^* \psi_m^{(b)}(z) \frac{e^{-\sqrt{\beta^2+q^2}|z-z_\ell|}}{\sqrt{\beta^2+q^2}}, \quad (140)$$

where the sum is over the considered bands. Equation (139) depends the detailed distribution of impurities, just as Eq. (101) depends on the detailed in-plane interface roughness  $\Delta_i(\mathbf{r})$ . By averaging over all possible distribution of impurities  $\mathbf{r}_i$  and  $\mathbf{r}_j$ , we can see that terms with  $\mathbf{r}_i = \mathbf{r}_j$  and  $\mathbf{q} = \mathbf{q}'$  dominate. Limiting the sum to  $\mathbf{r}_i = \mathbf{r}_j$  and going from sum over  $\mathbf{r}_i$  to integral gives

$$\begin{aligned} \sum_i e^{-i\mathbf{r}_i \cdot (\mathbf{q} - \mathbf{q}')} &\simeq \int d^2 r_i N_{2\text{D}} e^{-i\mathbf{r}_i \cdot (\mathbf{q} - \mathbf{q}')} \\ &= (2\pi)^2 N_{2\text{D}} \delta[\mathbf{q} - \mathbf{q}']. \end{aligned} \quad (141)$$

Inserting Eq. (141) into Eq. (139) gives

$$\begin{aligned} & \langle n_1, \mathbf{k}_1 | H_{\text{II}} | n_2, \mathbf{k}_2 \rangle \langle n_3, \mathbf{k}_2 | H_{\text{II}} | n_4, \mathbf{k}_4 \rangle \\ &= \frac{e^4}{16\pi^2\epsilon^2} N_{2D} \int d^2q \frac{1}{\beta^2 + q^2} (n_1 | n_2 )_{q,\ell}^{\text{II}} (n_2 | n_3 )_{q,\ell}^{\text{II}} \quad (142) \\ & \times \delta[\mathbf{k}_2 - (\mathbf{k}_1 - \mathbf{q})] \delta[\mathbf{k}_4 - (\mathbf{k}_2 + \mathbf{q})]. \end{aligned}$$

Inserting Eq. (142) into Eq. (138) and integrating over  $\mathbf{k}_4$  and  $\mathbf{k}_2$ , dropping the subscript on  $\mathbf{k}_1$  and shifting the integration variable  $\mathbf{q} \rightarrow -\mathbf{q} + \mathbf{k}$  gives

$$\begin{aligned} \mathcal{D}_{\text{II}}^{\text{in}} = & -\frac{e^4 N_{2D}}{16\pi\epsilon^2\hbar} \sum_{n_{1234}} \int d^2k \int d^2q \delta(E_q - \Delta_{n_4, n_3} - E_k) \\ & \times (n_1 | n_2 )_{|\mathbf{q}-\mathbf{k}|, \ell}^{\text{II}} (n_3 | n_4 )_{|\mathbf{q}-\mathbf{k}|, \ell}^{\text{II}} \rho_{n_2, n_3}^{E_q}. \quad (143) \end{aligned}$$

Calculating the scattering rates from Eq. (142) proceeds the same way as for interface roughness in Eqs. (104)–(108). The total rate due to ionized impurities is obtained by summing over all delta-doped sheets located at positions  $z = z_\ell$  with sheet densities  $N_{2D,\ell}$ . The result for the in-scattering rates is

$$\begin{aligned} \Gamma_{NMnmE_k}^{\pm, \text{in}, \text{II}} = & \frac{e^2 m_{\parallel}^*}{16\epsilon^2\hbar^3} u [E_k + \Delta_{N+M, N+M+m}] \sum_{\ell} \frac{N_{2D,\ell}}{2\pi} \\ & \times \left[ \int_0^{2\pi} d\theta (N|N+n)_{|\mathbf{q}-\mathbf{k}|, \ell}^{\text{II}} \right. \\ & \left. \times (N+M+m|N+M)_{|\mathbf{q}-\mathbf{k}|, \ell}^{\text{II}*} \right]_{E_q=\Delta_{N+M, N+M+m}+E_k} \quad (144) \end{aligned}$$

and the corresponding out-scattering rates are given by

$$\begin{aligned} \Gamma_{NMnmE_k}^{\pm, \text{out}, \text{II}} = & \frac{e^2 m_{\parallel}^*}{16\epsilon^2\hbar^3} u [E_k + \Delta_{N+n, N+M+m}] \sum_{\ell} \frac{N_{2D,\ell}}{2\pi} \\ & \times \left[ \int_0^{2\pi} d\theta (N+M|N+M+m)_{|\mathbf{q}-\mathbf{k}|, \ell}^{\text{II}} \right. \\ & \left. \times (N+M+m|N+n)_{|\mathbf{q}-\mathbf{k}|, \ell}^{\text{II}*} \right]_{E_q=\Delta_{N+n, N+M+m}+E_k}. \quad (145) \end{aligned}$$

The products involving the wavefunctions  $(n|m)_{|\mathbf{q}-\mathbf{k}|, \ell}^{\text{II}}$  are more complicated than in the case for both the phonon and interface roughness interaction. This is because the interaction potential for ionized impurities in Eq. (137) does not factor into the cross-plane and in-plane directions and the products involving eigenfunctions  $\psi_n(z)$  are coupled with the in-plane motion. To overcome this issue, we numerically calculate and store the quantity

$$F_{nm\ell}(b) = \frac{1}{b} \int \psi_n^*(z) \psi_m(z) e^{-b|z-z_\ell|} dz \quad (146)$$

using an evenly spaced mesh  $[b_{\min}, b_{\max}]$  for the relevant range of  $b$ . Typical values are  $b_{\min} = \beta$  and  $b_{\max} = 10 \text{ nm}^{-1}$ . We then perform the  $\theta$  integration numerically with the stored values of  $F_{nm\ell}(b)$  and convert between  $b$  and  $\theta$  using

$$b(\theta) = \sqrt{\frac{2m_{\parallel}^*}{\hbar^2}} \sqrt{E_D + E_k + E_q - 2\sqrt{E_k E_q} \cos(\theta)}. \quad (147)$$

with  $E_D = \hbar^2 \beta^2 / (2m_{\parallel}^*)$  the Debye energy.

## Appendix 6: Alloy scattering

Inserting the alloy interaction potential given by Eq. (101) into Eq. (100) gives the in-scattering part of the dissipator due to random alloy scattering

$$\begin{aligned} \mathcal{D}_{\text{alloy}}^{\text{in}} = & \frac{\pi}{\hbar} \sum_{n_{1234}} \int d^2k_{124} \delta[E(n_4, \mathbf{k}_4) - E(n_3, \mathbf{k}_2)] \rho_{n_2, n_3}^{E_{k_2}} \\ & \times \langle n_1, \mathbf{k}_1 | V_{\text{alloy}} | n_2, \mathbf{k}_2 \rangle \langle n_3, \mathbf{k}_2 | V_{\text{alloy}} | n_4, \mathbf{k}_4 \rangle \\ & \times | n_1, \mathbf{k}_1 \rangle \langle n_4, \mathbf{k}_4 | + \text{h.c..} \quad (148) \end{aligned}$$

The matrix elements involving the interaction potential can be simplified using

$$\begin{aligned} & \langle n_1, \mathbf{k}_1 | V_{\text{alloy}} | n_2, \mathbf{k}_2 \rangle \langle n_3, \mathbf{k}_2 | V_{\text{alloy}} | n_4, \mathbf{k}_4 \rangle \\ &= \sum_{ij} (n_1 | n_2 )_{\text{alloy}}(z_i) (n_3 | n_4 )_{\text{alloy}}(z_j) C_i^x C_j^x \\ & \times \langle \mathbf{k}_1 | \delta(\mathbf{r} - \mathbf{r}_i) | \mathbf{k}_2 \rangle \langle \mathbf{k}_2 | \delta(\mathbf{r} - \mathbf{r}_j) | \mathbf{k}_4 \rangle \\ &= \Omega_0 \sum_i (n_1 | n_2 )_{\text{alloy}}(z_i) (n_3 | n_4 )_{\text{alloy}}(z_i) x_i (1 - x_i) V_i^2 \\ & \times \langle \mathbf{k}_1 | \delta(\mathbf{r} - \mathbf{r}_i) | \mathbf{k}_2 \rangle \langle \mathbf{k}_2 | \delta(\mathbf{r} - \mathbf{r}_i) | \mathbf{k}_4 \rangle, \quad (149) \end{aligned}$$

where we have replaced the product  $C_i^x C_j^x$  with the correlation function in Eq. (113) and defined the alloy matrix element

$$(n|m)_{\text{alloy}}(z_i) = \sum_b [\psi_n^{(b)}(z_i)]^* \psi_m^{(b)}(z_i), \quad (150)$$

where the sum is over the considered bands. Going from sum to integral using  $\sum_i \rightarrow \frac{1}{\Omega_0} \int d^3R_i$  and following the same procedure as for interface roughness scattering in Eqs. (100)–(108) gives the scattering rates

$$\begin{aligned} \Gamma_{NMnmE_k}^{\pm, \text{in}, \text{alloy}} = & \frac{m_{\parallel}^* \Omega_0}{4\hbar^3} u [E_k + \Delta_{N+M, N+M+m}] \\ & \times \int dz (N|N+n)_{\text{alloy}}(z) (N+M+m|N+M)_{\text{alloy}}(z) \\ & \times x(z) (1 - x(z)) V(z)^2 \end{aligned}$$

and

$$\begin{aligned} \Gamma_{NMnmE_k}^{\pm,\text{out,alloy}} &= \frac{m_{\parallel}^* \Omega_0}{4\hbar^3} u [E_k + \Delta_{N+n,N+M+m}] \\ &\times \int dz (N+M|N+M+m)_{\text{alloy}}(z) \\ &(N+M+m|N+n)_{\text{alloy}}(z) \times x(z)(1-x(z))V(z)^2, \end{aligned}$$

where  $x(z)$  is the alloy fraction at position  $z$  and  $V(z)$  is the alloy strength parameter at position  $z$ , which we assume is the same for all bands.

## Material parameters

Table 1 lists all material specific parameters relating to the III-V binaries used in this work, as well as other parameters such as polar-optical-phonon energy and density. The temperature-dependent lattice constant is calculated using

$$a(T) = a(T = 300 \text{ K}) + \alpha_L(T - 300 \text{ K}), \quad (151)$$

with  $a(T = 300 \text{ K})$  the lattice constant at 300 K and the thermal expansion coefficient  $\alpha_L$  given in table 1. In order to calculate the temperature-dependent  $\Gamma$ -valley bandgap, we use the empirical Varshni functional form [85]

$$E_{\text{gap}}(T) = E_{\text{gap}}(T = 0) - \frac{\alpha^{\Gamma} T^2}{T + \beta^{\Gamma}}, \quad (152)$$

where  $E_{\text{gap}}(T = 0)$  is the zero-temperature  $\Gamma$ -valley bandgap and  $\alpha^{\Gamma}$  and  $\beta^{\Gamma}$  are the Varshni parameters given in table 1. All parameters relating to band structure in table 1

are taken from Ref. [73] with the exception of Kane energy for GaAs and InAs, which were chosen to give room-temperature effective electron masses of  $0.063m_0$  and  $0.023m_0$ , respectively ( $m_0$  is the free electron mass). Note that the conduction-band effective mass can be written in terms of the parameters in table 1

$$\frac{m_0}{m_c^*} = (1 + 2F) + \frac{E_p(E_{\text{gap}} + 2\Delta_{\text{so}}/3)}{E_{\text{gap}}(E_{\text{gap}} + \Delta_{\text{so}})}. \quad (153)$$

Material parameters such as Kane energy for ternary alloys ( $A_{1-x}B_xC$ ) as a function of alloy composition  $x$  are obtained using a second-order interpolation formula on the form [86]

$$E_p(A_{1-x}B_xC) = (1 - x)E_p(A) + xE_p(B) - x(1 - x)C_b, \quad (154)$$

where  $E_p(A)$  and  $E_p(B)$  are the Kane energies of binary AC and BC, respectively, and the bowing parameter  $C_b$  accounts for deviation from linear interpolation. Generalization to other material parameters is straightforward. Exceptions to Eq. (154) are interpolations for the Kane energy and F parameters of the  $\text{In}_{1-x}\text{Ga}_x\text{As}$  alloy, which contain a third-order term due to the sensitivity of energy spacing to the effective mass of the well material. This third-order term is described in more detail later in this appendix.

Nonzero bowing parameters for the  $\text{In}_{1-x}\text{Ga}_x\text{As}$  and  $\text{In}_{1-x}\text{Al}_x\text{As}$  alloys are given in table 2. All parameters are

**Table 1** Various material parameters for GaAs, AlAs and InAs

Parameter	Symbol	GaAs	AlAs	InAs
Lattice constant at 300 K (Å)	$a(T = 300 \text{ K})$	5.65325	5.6611	6.0583
Thermal expansion coefficient ( $10^{-5}\text{\AA/K}$ )	$\alpha_L$	3.88	2.90	2.74
Bandgap ( $\Gamma$ valley) at 0 K (eV)	$E_{\text{gap}}(T = 0)$	1.519	3.099	0.417
Varshni parameter (meV/K)	$\alpha^{\Gamma}$	0.5405	0.885	0.276
Varshni parameter (K)	$\beta^{\Gamma}$	204	530	93
Spin-orbit splitting (eV)	$\Delta_{\text{so}}$	0.341	0.28	0.39
Luttinger parameter	$\gamma_1^L$	6.98	3.76	20.0
Luttinger parameter	$\gamma_2^L$	2.06	0.82	8.50
Kane energy (eV)	$E_p$	28.6	21.1	20.7
F parameter	$F$	-1.94	-0.48	-2.9
Valence-band offset (eV)	VBO	-0.80	-1.33	-0.59
Conduction band deformation potential (eV)	$a_c$	-7.17	-5.64	-5.08
Valence band deformation potential (eV)	$a_v$	-1.16	-2.47	-1.00
Shear deformation potential (eV)	$b$	-2.00	-2.30	-1.80
Elastic constant (GPa)	$c_{11}$	122.1	125.0	83.3
Elastic constant (GPa)	$c_{12}$	56.6	53.4	45.3
Polar-optical phonon energy (meV)	$E_{\text{POP}}$	35	50	30
Mass density (g/cm <sup>3</sup> )	$\rho$	5.32	3.76	5.68
Low-frequency relative dielectric constant	$\epsilon_s^0$	12.90	10.06	15.15
High-frequency relative dielectric constant	$\epsilon_s^{\infty}$	10.89	8.16	12.30

obtained from Ref. [73], except for the Kane energy and F parameter in  $\text{In}_{1-x}\text{Ga}_x\text{As}$ , which in this work were chosen to give a room-temperature electron effective mass of  $0.041m_0$  for  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  (lattice-matched to InP). Note that when interpolating Luttinger parameters, the light holes ( $m_{lh}^*$ ) and heavy hole ( $m_{hh}^*$ ) masses should be interpolated using the bowing parameters in table 2 and corresponding Luttinger parameters calculated afterwards using

$$\begin{aligned} \frac{m_0}{m_{lh}^*} &= \gamma_1^L + 2\gamma_2^L \\ \frac{m_0}{m_{hh}^*} &= \gamma_1^L - 2\gamma_2^L. \end{aligned} \quad (155)$$

As shown in Sect. 4.1, the material parameters presented in tables 1 and 2 give results in very good agreement with experiment as well as theory based on NEGF for lattice-matched designs. However, for the strain-balanced design in Sect. 4.2, the peak gain wavelength was overestimated by  $0.2\text{ }\mu\text{m}$ , with an associated shift to lower electric fields in the current density vs electric field curve. We attribute this disagreement to underestimation of the electron mass in the well material (lower mass means smaller energy spacing) and addressed the issue by adding a cubic term

$$-x(1-x)(x-0.47)D \quad (156)$$

to the alloy interpolation formula [Eq. (154)] for Kane energy and F parameter in  $\text{In}_{1-x}\text{Ga}_x\text{As}$ . This functional form was chosen to not change the lattice-matched results for  $x = 0.47$ , or the binary results ( $x = 0$  and  $x = 1$ ). Values of  $D = 32\text{ eV}$  for Kane energy and  $D = 15$  for the F parameter were found to reproduce experimental results with an energy spacing of  $270\text{ meV}$  (corresponding to  $\lambda = 4.6\text{ }\mu\text{m}$ ) between the upper and lower lasing states. These values of  $D$  might seem excessively high. However,  $D$  is multiplied by 3 numbers that have magnitudes smaller than 1.

We will end this Appendix with a discussion of the parameters related to interface-roughness scattering. We used the same interface-roughness correlation length

**Table 2** Nonzero bowing parameters for the  $\text{In}_{1-x}\text{Ga}_x\text{As}$  and  $\text{In}_{1-x}\text{Al}_x\text{As}$  alloys

Parameter	$\text{In}_{1-x}\text{Ga}_x\text{As}$	$\text{In}_{1-x}\text{Al}_x\text{As}$
$E_{\text{gap}}^F\text{ (eV)}$	0.477	0.70
$\Delta_{\text{so}}\text{ (eV)}$	0.15	0.15
$m_{hh}^*$	-0.145	0.00
$m_{lh}^*$	0.0202	0.00
$E_p\text{ (eV)}$	3.20	-4.81
$F$	-1.00	-4.44
VBO (eV)	-0.38	-0.64
$a_c\text{ (eV)}$	2.61	-1.40

$\Lambda = 9\text{ nm}$  as in Refs. [32, 34]. We estimate the RMS value for interface roughness ( $\Delta$ ) by using the fact that interface-roughness scattering is the main broadening mechanism of the gain spectrum. [1] We followed the procedure in Ref. [32] and fix the correlation length at  $9\text{ nm}$  and vary  $\Delta$  such that the gain spectra produces a peak with a FWHM of  $25\text{ meV}$ , matching experimental results for electroluminescence spectra at a field strength of  $53\text{ kV/cm}$ . [52] We note that the resulting value of  $\Delta = 0.08\text{ nm}$  is smaller than the value of  $\Delta = 0.1\text{ nm}$  used in Ref. [34] to simulate the same device. We attribute the difference to the different electron Hamiltonian used in this work (three-band  $\mathbf{k} \cdot \mathbf{p}$ ), resulting in higher amplitude of eigenfunctions at material interfaces.

## References

1. Faist, J., Capasso, F., Sivco, D.L., Sirtori, C., Hutchinson, A.L., Cho, A.Y.: Quantum cascade laser. *Science* **264**(5158), 553–556 (1994)
2. Vitiello, M.S., Scalari, G., Williams, B., De Natale, P.: Quantum cascade lasers: 20 years of challenges. *Opt. Express* **23**(4), 5167–5182 (2015)
3. Capasso, F.: High-performance midinfrared quantum cascade lasers. *Opt. Eng.* **49**(11), 111102 (2010)
4. Yao, Yu., Hoffman, A.J., Gmachl, C.F.: Mid-infrared quantum cascade lasers. *Nat. Photon* **6**(7), 432 (2012)
5. Razeghi, M., Bandyopadhyay, N., Bai, Y., Quanyong, L., Slivken, S.: Recent advances in mid infrared ( $3\text{--}5\text{ }\mu\text{m}$ ) quantum cascade lasers. *Opt. Mater. Express* **3**(11), 1872–1884 (2013)
6. Cur, R.F., Capasso, F., Gmachl, C., Kosterev, A.A., McManus, B., Lewicki, R., Pusharsky, M., Wysocki, G., Tittel, F.K.: Quantum cascade lasers in chemical physics. *Chem. Phys. Lett.* **487**(1–3), 1–18 (2010)
7. Bartalini, S., Vitiello, M.S., De Natale, P.: Quantum cascade lasers: a versatile source for precise measurements in the mid/far-infrared range. *Meas. Sci. Technol.* **25**(1), 012001 (2013)
8. Dupont, E., Fathololoumi, S., Liu, H.C.: Simplified density-matrix model applied to three-well terahertz quantum cascade lasers. *Phys. Rev. B* **81**(20), 205311 (2010)
9. Jirauschek, C., Kubis, T.: Modeling techniques for quantum cascade lasers. *Appl. Phys. Rev.* **1**(1), 011307 (2014)
10. Indjin, D., Harrison, P., Kelsall, R.W., Ikonić, Z.: Self-consistent scattering theory of transport and output characteristics of quantum cascade lasers. *J. Appl. Phys.* **91**(11), 9019–9026 (2002)
11. Indjin, D., Harrison, P., Kelsall, R.W., Ikonić, Z.: Influence of leakage current on temperature performance of gaas/algaa quantum cascade lasers. *Appl. Phys. Lett.* **81**(3), 400–402 (2002)
12. Mirčetić, A., Indjin, D., Ikonić, Z., Harrison, P., Milanović, V., Kelsall, R.W.: Towards automated design of quantum cascade lasers. *J. Appl. Phys.* **97**(8), 084506 (2005)
13. Wang, X.-G., Grillot, F., Wang, C.: Rate equation modeling of the frequency noise and the intrinsic spectral linewidth in quantum cascade lasers. *Opt. Express* **26**(3), 2325–2334 (2018)
14. Rita Claudia Iotti and Fausto Rossi: Nature of charge transport in quantum-cascade lasers. *Phys. Rev. Lett.* **87**(14), 146603 (2001)
15. Callebaut, H., Kumar, S., Williams, B.S., Hu, Q., Reno, J.L.: Importance of electron-impurity scattering for electron transport in terahertz quantum-cascade lasers. *Appl. Phys. Lett.* **84**(5), 645–647 (2004)

16. Gao, X., Botez, D., Knezevic, I.: X-valley leakage in gas-based midinfrared quantum cascade lasers: a Monte Carlo study. *J. Appl. Phys.* **101**(6), 063101 (2007)
17. Shi, Y.B., Knezevic, I.: Nonequilibrium phonon effects in midinfrared quantum cascade lasers. *J. Appl. Phys.* **116**(12), 123105 (2014)
18. Borowik, P., Thobel, J.-L., Adamowicz, L.: Monte carlo modeling applied to studies of quantum cascade lasers. *Opt. Quant. Electron.* **49**(3), 96 (2017)
19. Willenberg, H., Döhler, G.H., Faist, J.: Intersubband gain in a bloch oscillator and quantum cascade laser. *Phys. Rev. B* **67**(8), 085315 (2003)
20. Kumar, S., Qing, H.: Coherence of resonant-tunneling transport in terahertz quantum-cascade lasers. *Phys. Rev. B* **80**(24), 245316 (2009)
21. Weber, C., Wacker, A., Knorr, A.: Density-matrix theory of the optical dynamics and transport in quantum cascade structures: the role of coherence. *Phys. Rev. B* **79**(16), 165322 (2009)
22. Terazzi, R., Faist, J.: A density matrix model of transport and radiation in quantum cascade lasers. *New J. Phys.* **12**(3), 033045 (2010)
23. Burnett, B.A., Pan, A., Chui, C.O., Williams, B.S.: Robust density matrix simulation of terahertz quantum cascade lasers. *IEEE Trans. Terahertz Sci. Technol.* **8**(5), 492–501 (2018)
24. Pan, A., Burnett, B.A., Chui, C.n.: Density matrix modeling of quantum cascade lasers without an artificially localized basis: a generalized scattering approach. *Phys. Rev. B* **96**(8), 085308 (2017)
25. Demić, A., Grier, A., Ikonić, Z., Valavanis, A., Evans, C.A., Mohandas, R., Li, L., Linfield, E.H., Davies, A.G., Indjin, D.: Infinite-period density-matrix model for terahertz-frequency quantum cascade lasers. *IEEE Trans. Terahertz Sci. Technol.* **7**(4), 368–377 (2017)
26. Jirauschek, C.: Density matrix monte carlo modeling of quantum cascade lasers. *J. Appl. Phys.* **122**(13), 133105 (2017)
27. Riesch, M., Tchipev, N., Senninger, S., Bungartz, H.-J., Jirauschek, C.: Performance evaluation of numerical methods for the Maxwell–Liouville–on neumann equations. *Opt. Quant. Electron.* **50**(2), 112 (2018)
28. Riesch, M., Jirauschek, C.: Analyzing the positivity preservation of numerical methods for the Liouville–von Neumann equation. *J. Comput. Phys.* **390**, 290–296 (2019)
29. Demic, A., Ikonic, Z., Kelsall, R.W., Indjin, D.: Density matrix superoperator for periodic quantum systems and its application to quantum cascade laser structures. *AIP Adv.* **9**(9), 095019 (2019)
30. Jirauschek, C., Riesch, M., Tzenov, P.: Optoelectronic device simulations based on macroscopic Maxwell–Bloch equations. *Adv. Theory Simul.* **2**(8), 1900018 (2019)
31. Lee, S.-C., Wacker, A.: Nonequilibrium Green’s function theory for transport and gain properties of quantum cascade structures. *Phys. Rev. B* **66**(24), 245314 (2002)
32. Bugajski, M., Gutowski, P., Karbownik, P., Kolek, A., Hałdaś, G., Pierściński, K., Pierścińska, D., Kubacka-Traczyk, J., Sankowska, I., Trajnerowicz, A., et al.: Mid-ir quantum cascade lasers: device technology and non-equilibrium green’s function modeling of electro-optical characteristics. *Phys. Status Solidi B* **251**(6), 1144–1157 (2014)
33. Wacker, A., Lindskog, M., Winge, D.O.: Nonequilibrium green’s function model for simulation of quantum cascade laser devices under operating conditions. *IEEE J. Sel. Top. Quantum Electron.* **19**(5), 1–11 (2013)
34. Lindskog, M., Wolf, J.M., Trinite, V., Liverini, V., Faist, Jérôme, Maisons,G., Carras, M., Aidam, R., Ostendorf, R., Wacker, A.: Comparative analysis of quantum cascade laser modeling based on Density matrices and non-equilibrium green’s functions. *Appl. Phys. Lett.*, **105**(10), 103106, (2014)
35. Kolek, A., Hałdaś, G., Bugajski, M., Pierściński, K., Gutowski, P.: Impact of injector doping on threshold current of mid-infrared quantum cascade laser-non-equilibrium green’s function analysis. *IEEE J. Sel. Top. Quantum Electron.* **21**(1), 124–133 (2014)
36. Hałdaś, G.: Implementation of non-uniform mesh in non-equilibrium green’s function simulations of quantum cascade lasers. *J. Comput. Electron.* 1–7 (2019)
37. Kolek, A.: Implementation of light-matter interaction in negf simulations of qcl. *Opt. Quant. Electron.* **51**(6), 171 (2019)
38. Grange, T., Stark, D., Scalari, G., Faist, J., Persichetti, Luca, Di Gaspare, Luciana, De Seta, Monica, Ortolani, Michele, Paul, Douglas J., Capellini, Giovanni: et al. Room temperature operation of n-type ge/sige terahertz quantum cascade lasers predicted by non-equilibrium Green’s functions. *Appl. Phys. Lett.* **114**(11), 111102 (2019)
39. Callebaut, H., Qing, H.: Importance of coherence for electron transport in terahertz quantum cascade lasers. *J. Appl. Phys.* **98**(10), 104505 (2005)
40. Jonasson, O., Mei, Song, Karimi, Farhad, Kirch, J., Botez, Dan, Mawst, Luke, Knezevic, Irena: Quantum transport simulation of high-power 4.6- $\mu\text{m}$  quantum cascade lasers. In: Photonics, volume 3, page 38. Multidisciplinary Digital Publishing Institute (2016)
41. Iotti, R.C., Rossi, F.: Microscopic theory of semiconductor-based optoelectronic devices. *Rep. Prog. Phys.* **68**(11), 2533 (2005)
42. Jonasson, O., Karimi, F., Knezevic, I.: Partially coherent electron transport in terahertz quantum cascade lasers based on a markovian master equation for the density matrix. *J. Comput. Electron.* **15**(4), 1192–1205 (2016)
43. Lindblad, G.: On the generators of quantum dynamical semigroups. *Commun. Math. Phys.* **48**(2), 119 (1976)
44. Heinz-Peter, B., Francesco, P., et al.: The Theory of Open Quantum Systems. Oxford University Press on Demand, Oxford
45. Knezevic, I., Novakovic, B.: Time-dependent transport in open systems based on quantum master equations. *J. Comput. Electron.* **12**(3), 363–374 (2013)
46. Karimi, F.: Quantum transport theory on optical and plasmonic properties of nanomaterials. PhD thesis, University of Wisconsin–Madison (2017)
47. Karimi, F., Davoodi, A.H., Knezevic, I.: Dielectric function and plasmons in graphene: a self-consistent-field calculation within a Markovian master equation formalism. *Phys. Rev. B* **93**, 205421 (2016)
48. Fischetti, M.V.: Master-equation approach to the study of electronic transport in small semiconductor devices. *Phys. Rev. B* **59**(7), 4901 (1999)
49. Haug, H., Koch, S.W.: Quantum Theory of the Optical and Electronic Properties of Semiconductors, 5h edn. World Scientific, Singapore (2009)
50. Calecki, D., Palmier, J.F., Chomette, A.: Hopping conduction in multiquantum well structures. *J. Phys. C: Solid state* **17**(28), 5017 (1984)
51. Gelmont, B., Gorfinkel, V., Luryi, S.: Theory of the spectral line shape and gain in quantum wells with intersubband transitions. *Appl. Phys. Lett.* **68**(16), 2171–2173 (1996)
52. Jirauschek, C., Lugli, P.: Monte-Carlo-based spectral gain analysis for terahertz quantum cascade lasers. *J. Appl. Phys.* **105**(12), 123102 (2009)
53. Saad, Y.: Iterative Methods for Sparse Linear Systems, vol. 82. SIAM, Philadelphia (2003)
54. Bismuto, A., Terazzi, R., Beck, M., Faist, J.: Electrically tunable, high performance quantum cascade laser. *Appl. Phys. Lett.* **96**(14), 141105 (2010)
55. Mátyás, A., Lugli, P., Jirauschek, C.: Photon-induced carrier transport in high efficiency midinfrared quantum cascade lasers. *J. Appl. Phys.* **110**(1), 013108 (2011)

56. Claudia Iotti, R., Rossi, F.: Carrier thermalization versus phonon-assisted relaxation in quantum-cascade lasers: A Monte Carlo approach. *Appl. Phys. Lett.* **78**(19), 2902–2904 (2001)
57. Spagnolo, V., Scamarcio, G., Page, H., Sirtori, C.: Simultaneous measurement of the electronic and lattice temperatures in gaas/al 0.45 ga 0.55 as quantum-cascade lasers: influence on the optical performance. *Appl. Phys. Lett.* **84**(18), 3690–3692 (2004)
58. Page, H., Becker, C., Robertson, A., Glastre, G., Ortiz, V., Sirtori, C.: 300 K operation of a gaas-based quantum-cascade laser at  $\lambda$  9  $\mu\text{m}$ . *Appl. Phys. Lett.* **78**(22), 3529–3531 (2001)
59. Lundstrom, M.: Fundamentals of Carrier Transport. Cambridge University Press, Cambridge (2009)
60. Bai, Y., Darvish, S.R., Slivken, S., Zhang, W., Evans, A., Nguyen, J., Razeghi, M.: Room temperature continuous wave operation of quantum cascade lasers with watt-level optical power. *Appl. Phys. Lett.* **92**(10), 101105 (2008)
61. Evans, A., Darvish, S.R., Slivken, S., Nguyen, J., Bai, Y., Razeghi, M.: Buried heterostructure quantum cascade lasers with high continuous-wave wall plug efficiency. *Appl. Phys. Lett.* **91**(7), 071101 (2007)
62. Iotti, R.C., Rossi, F.: Microscopic theory of quantum-cascade lasers. *Semicond. Sci. Technol.* **19**(4), S323 (2004)
63. Bastard, G.: Wave Mechanics Applied to Semiconductor Heterostructures (1990)
64. Chuang, S.L., Chuang, S.L.: Physics of Optoelectronic Devices (1995)
65. Nelson, D.F., Miller, R.C., Kleinman, D.A.: Band nonparabolicity effects in semiconductor quantum wells. *Phys. Rev. B* **35**(14), 7770 (1987)
66. Dupont, E., Fathololoumi, S., Wasilewski, Z.R., Aers, G., Laframboise, S.R., Lindskog, M., Razavipour, S.G., Wacker, A., Ban, D., Liu, H.C.: A phonon scattering assisted injection and extraction based terahertz quantum cascade laser. *J. Appl. Phys.* **111**(7), 073111 (2012)
67. Terazzi, R.L.: Transport in quantum cascade lasers. Ph.D. thesis, ETH Zurich (2012)
68. Kolek, A., Hałdaś, G., Bugajski, M.: Nonthermal carrier distributions in the subbands of 2-phonon resonance mid-infrared quantum cascade laser. *Appl. Phys. Lett.* **101**(6), 061110 (2012)
69. Ekenberg, U.: Nonparabolicity effects in a quantum well: sublevel shift, parallel mass, and landau levels. *Phys. Rev. B* **40**(11), 7714 (1989)
70. Sirtori, C., Capasso, F., Faist, J., Scandolo, S.: Nonparabolicity and a sum rule associated with bound-to-bound and bound-to-continuum intersubband transitions in quantum wells. *Phys. Rev. B* **50**(12), 8663 (1994)
71. Bahder, T.B.: Eight-band k.p model of strained zinc-blende crystals. *Phys. Rev. B* **41**(17), 11992 (1990)
72. Kane, E.O.: Semiconductors and Semimetals, vol. 1. Academic Press, London, p. 75 (1966)
73. Pidgeon, C.R., Brown, R.N.: Interband magneto-absorption and faraday rotation in insb. *Phys. Rev.* **146**(2), 575 (1966)
74. Luttinger, J.M., Kohn, W.: Motion of electrons and holes in perturbed periodic fields. *Phys. Rev.* **97**(4), 869 (1955)
75. Vurgaftman, I., Meyer, J.R., Ram-Mohan, R.L.: Band parameters for iii–v compound semiconductors and their alloys. *J. Appl. Phys.* **89**(11), 5815–5875 (2001)
76. Bir, G.L., Pikus, G.E.: Symmetry and strain-induced effects in semiconductors (1974)
77. Voon, L.Y., Willatzen, L.C.: The kp method: electronic properties of semiconductors. Springer, Berlin (2009)
78. Fröhlich, H.: Theory of electrical breakdown in ionic crystals. *Proc. R. Soc. Lond. A* **160**(901), 230–241 (1937)
79. Kubis, T., Yeh, C., Vogl, P.: Quantum theory of transport and optical gain in quantum cascade lasers. *Phys. Status Solidi C* **5**(1), 232–235 (2008)
80. Chiu, Y.T., Dikmelik, Y., Liu, P.Q., Aung, N.L., Khurgin, J.B., Gmachl, C.F.: Importance of interface roughness induced inter-subband scattering in mid-infrared quantum cascade lasers. *Appl. Phys. Lett.* **101**(17), 171117 (2012)
81. Ferry, D.K.: Semiconductors, 2053–2563 (2013)
82. Roblin, P., Potter, R.C., Fathimulla, A.: Interface roughness scattering in alas/ingaas resonant tunneling diodes with an inas subwell. *J. Appl. Phys.* **79**(5), 2502–2508 (1996)
83. Ando, T.: Self-consistent results for a gaas/al x ga1-x as heterojunciton. II. Low temperature mobility. *J. Phys. Soc. Jpn.* **51**(12), 3900–3907 (1982)
84. Gradshteyn, I.S., Ryzhik, I.M.: Table of integrals, series and products 7th edn. In: Jeffrey, A., Zwilinger, D. (eds.) New York: Academic (2007)
85. Varshni, Y.P.: Temperature dependence of the energy gap in semiconductors. *Physica* **34**(1), 149–154 (1967)
86. Van Vechten, J.A., Bergstresser, T.K.: Electronic structures of semiconductor alloys. *Phys. Rev. B* **1**(8), 3351 (1970)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.