

Partially coherent electron transport in terahertz quantum cascade lasers based on a Markovian master equation for the density matrix

O. Jonasson¹  · F. Karimi¹ · I. Knezevic¹ 

© Springer Science+Business Media New York 2016

Abstract We derive a Markovian master equation for the single-electron density matrix, applicable to quantum cascade lasers (QCLs). The equation conserves the positivity of the density matrix, includes off-diagonal elements (coherences) as well as in-plane dynamics, and accounts for electron scattering with phonons and impurities. We use the model to simulate a terahertz-frequency QCL, and compare the results with both experiment and simulation via nonequilibrium Green's functions (NEGF). We obtain very good agreement with both experiment and NEGF when the QCL is biased for optimal lasing. For the considered device, we show that the magnitude of coherences can be a significant fraction of the diagonal matrix elements, which demonstrates their importance when describing THz QCLs. We show that the in-plane energy distribution can deviate far from a heated Maxwellian distribution, which suggests that the assumption of thermalized subbands in simplified density-matrix models is inadequate. We also show that the current density and subband occupations relax toward their steady-state values on very different time scales.

Keywords QCL · Superlattice · Quantum transport · Dissipation · Density matrix · Phonons · Terahertz

1 Introduction

Quantum cascade lasers (QCLs) are semiconductor heterostructures that operate based on quantum confinement and tunneling. Population inversion between quasi-bound lasing states is achieved through precise engineering of material composition and layer widths [1]. Numerical simulations play an important role in the design of QCLs [2,3]. For this purpose, a range of theoretical models have been employed, including semiclassical [4–7] and quantum-transport techniques based on the density-matrix formalism [2,8–11] or nonequilibrium Green's functions (NEGF) [12]. Semiclassical approaches are appealing owing to their low computational requirements. They go beyond the effective-mass approximation [6] and can explore phenomena such as nonequilibrium phonons [7]. However, semiclassical models can provide inadequate descriptions of QCLs working in the THz range, where the role of coherence cannot be ignored [9,13].

In order to maximize the performance of THz QCLs, optimization methods such as the genetic algorithm have been used, where the simulation converges on a layer structure that maximizes the gain of the device [14]. These simulations require repeated calculations of device performance for a large number of parameters, so computational efficiency plays an important role. This fact makes density-matrix-based approaches advantageous over the relatively high computational burden of NEGF [15]. However, common density-matrix-based approaches have two significant drawbacks. One is a common assumption of thermalized subbands, where the electron temperature is either an input parameter [9,15] or determined using an energy-balance method [16]. This approximation may not be warranted, because QCLs operate far from equilibrium, so the in-plane energy distribution can (as will be shown later in this work)

✉ O. Jonasson
ojonasson@wisc.edu

F. Karimi
karimi2@wisc.edu

I. Knezevic
irena.knezevic@wisc.edu

¹ University of Wisconsin-Madison, Madison, WI 53706, USA

deviate far from a heated thermal distribution (Maxwellian or Fermi–Dirac), which makes electron temperature an ill-defined quantity. The second drawback is phenomenological treatment of dephasing across thick injection barriers: transport is treated semiclassically within a single stage of a QCL (or a part of a single stage), while interstage dynamics are treated quantum mechanically using phenomenological dephasing times [9, 11, 13, 15].

In this work, we propose a computationally efficient density-matrix model based on a rigorously derived Markovian master equation, where neither of the two aforementioned simplifications are made. The Markovian master equation conserves the positivity of the density matrix, includes off-diagonal matrix elements as well as full in-plane dynamics and time dependence, and accounts for the relevant scattering mechanisms with phonons and impurities. We apply the model to a terahertz QCL proposed in Ref. [14]. With the QCL biased for lasing, we obtain very good agreement with experiment, as well as with theoretical results based on NEGF. We show that the magnitude of the off-diagonal elements of the density matrix (which we will refer to as *coherences*) can be a significant fraction of the diagonal values, demonstrating the importance of including coherences when describing THz QCLs. We show that significant electron heating takes place, with subband electron distributions over in-plane kinetic energy deviating far from thermal. Last, we provide time-resolved results, giving insight into the response of the device to a suddenly applied bias, thus revealing the different time scales involved.

This paper is organized into five sections and an appendix. In Sect. 2, we derive a Markovian master equation for the single-electron density matrix that is applicable to electron transport in QCLs. In Sect. 3, we describe the numerical solution method. Results for a THz QCL are given in Sect. 4, along with a comparison to NEGF and experiment. Section 5 contains concluding remarks.

2 Derivation of the master equation

In the following, we will denote three-dimensional (3D) vectors by uppercase letters and two-dimensional (2D) vectors by lowercase letters; for example, $\mathbf{Q} = (Q_x, Q_y, Q_z)$ and $\mathbf{k} = (k_x, k_y)$. Transport is in the z -direction (cross-plane) and translational invariance in the x - y plane (in-plane) is assumed. $\mathbf{Q} + \mathbf{k}$ should be understood as $(Q_x + k_x, Q_y + k_y, Q_z)$.

The total Hamiltonian of an open electronic system, describing the behavior of electrons interacting with a dissipative phonon bath, can be written as

$$\hat{H} = \hat{H}_0 + \hat{H}_{e\text{-ph}} + \hat{H}_{\text{ph}}. \tag{1}$$

\hat{H}_0 is the unperturbed Hamiltonian of electrons, including the kinetic and potential electronic terms, and \hat{H}_{ph} denotes

the Hamiltonian of the free phonon bath. The interaction Hamiltonian between electrons and phonons is included in $\hat{H}_{e\text{-ph}}$. We use a Frölich-type Hamiltonian to describe the interaction of a single electron with a phonon bath [17]:

$$\hat{H}_{e\text{-ph}} = \frac{1}{(2\pi)^3} \sum_g \int d^3Q \mathcal{M}_g(\mathbf{Q}) \left(b_{g,\mathbf{Q}} e^{i\mathbf{Q}\cdot\hat{\mathbf{R}}} - b_{g,\mathbf{Q}}^\dagger e^{-i\mathbf{Q}\cdot\hat{\mathbf{R}}} \right). \tag{2}$$

Here, $b_{g,\mathbf{Q}}^\dagger$ ($b_{g,\mathbf{Q}}$) is the phonon creation (annihilation) operator for a phonon in branch g with wave vector \mathbf{Q} and $\mathcal{M}_g(\mathbf{Q})$ is the associated scattering matrix element. Note that we have assumed the phonon wave vectors are closely spaced to warrant integration over \mathbf{Q} . The equation of motion for the statistical operator ($\hat{\rho}$) in the interaction picture is

$$\begin{aligned} \frac{d}{dt} \tilde{\rho}(t) &= -\frac{i}{\hbar} \left[\tilde{H}_{e\text{-ph}}(t), \tilde{\rho}(t) \right], \\ \tilde{\rho}(t) &= \tilde{\rho}(0) - \frac{i}{\hbar} \int_0^t \left[\tilde{H}_{e\text{-ph}}(t'), \tilde{\rho}(t') \right] dt'. \end{aligned} \tag{3}$$

The tilde symbol denotes that the operators are in the interaction picture, i.e., $\tilde{O}(t) = e^{\frac{i}{\hbar}(\hat{H}_0 + \hat{H}_{\text{ph}})t} O e^{-\frac{i}{\hbar}(\hat{H}_0 + \hat{H}_{\text{ph}})t}$. We assume that the interaction of the electron and phonons only negligibly affects the density matrix of the phonon reservoir (Born approximation): thus, the density matrix of the total system may be represented as a tensor product: $\tilde{\rho}(t) = \tilde{\rho}_e(t) \otimes \tilde{\rho}_{\text{ph}}$ [18, 19]. We also assume the interaction strength is sufficiently high to treat the system as memoryless (Markov approximation), i.e., the evolution of the density matrix only depends on its present state. Now, we put the integral form in Eq. (3) in the right-hand side of the differential form, then we apply the Born and Markov approximations, and finally we take the trace over the phonon reservoir. Then, the equation of motion reads as follows:

$$\begin{aligned} \frac{d}{dt} \tilde{\rho}_e(t) &= -\frac{i}{\hbar} \text{tr}_{\text{ph}} \left\{ \left[\tilde{H}_{e\text{-ph}}(t), \tilde{\rho}_e(0) \otimes \tilde{\rho}_{\text{ph}} \right] \right\} \\ &\quad - \frac{1}{\hbar^2} \int_0^\infty ds \text{tr}_{\text{ph}} \left\{ \left[\tilde{H}_{e\text{-ph}}(t), \left[\tilde{H}_{e\text{-ph}}(t-s), \tilde{\rho}_e(t) \otimes \tilde{\rho}_{\text{ph}} \right] \right] \right\}. \end{aligned} \tag{4}$$

In order to remove the temporal dependence of the interaction Hamiltonian, we switch back to the Schrödinger picture and use $\text{tr}_{\text{ph}} \left\{ \hat{H}_{e\text{-ph}} \hat{\rho}_{\text{ph}} \right\} = 0$, which gives

$$\begin{aligned} \frac{d\hat{\rho}_e(t)}{dt} &= -\frac{i}{\hbar} [\hat{H}_0, \hat{\rho}_e(t)] - \frac{1}{\hbar^2} \int_0^\infty ds \\ &\quad \times \text{tr}_{\text{ph}} \left\{ \left[\hat{H}_{e\text{-ph}}, \left[e^{-i(\hat{H}_0 + \hat{H}_{\text{ph}})s/\hbar} \hat{H}_{e\text{-ph}} e^{i(\hat{H}_0 + \hat{H}_{\text{ph}})s/\hbar}, \hat{\rho}_e(t) \otimes \hat{\rho}_{\text{ph}} \right] \right] \right\}. \end{aligned} \tag{5}$$

We will refer to the second term on the right-hand side of the above equation as $\hat{\mathcal{D}}$, the *dissipation superoperator* or

the dissipator, acting on the density matrix. The equation of motion for the reduced single-electron density operator $\hat{\rho}_e$ can then be written as

$$\frac{\partial \hat{\rho}_e}{\partial t} = -\frac{i}{\hbar}[\hat{H}_0, \hat{\rho}_e] + \hat{D}(\hat{\rho}_e), \tag{6}$$

where \hat{D} contains the effect of dissipation due to interactions with phonons; spatially random static disorder can also be included in \hat{D} [20–22]. By tracing over the phonon degree of freedom in (5) and expanding the commutators, \hat{D} can be grouped into eight terms, containing four hermitian conjugate pairs. Two terms correspond to emission and two to absorption. In order to keep the equations compact, calculations will only be shown explicitly for the emission terms. Using this simplification, we can write

$$\begin{aligned} \hat{D}(\hat{\rho}_e) = & \frac{1}{\hbar^2} \frac{1}{(2\pi)^3} \int d^3 Q \int_0^\infty ds \{ \\ & - \mathcal{W}_g^{\text{em}}(\mathbf{Q}) e^{-iE_g s/\hbar} e^{-i\mathbf{Q}\cdot\hat{\mathbf{R}}} e^{-i\hat{H}_0 s/\hbar} e^{i\mathbf{Q}\cdot\hat{\mathbf{R}}} e^{i\hat{H}_0 s/\hbar} \hat{\rho}_e \\ & + \mathcal{W}_g^{\text{em}}(\mathbf{Q}) e^{+iE_g s/\hbar} e^{i\mathbf{Q}\cdot\hat{\mathbf{R}}} \hat{\rho}_e e^{-i\hat{H}_0 s/\hbar} e^{-i\mathbf{Q}\cdot\hat{\mathbf{R}}} e^{i\hat{H}_0 s/\hbar} \\ & + \text{h.c.} + \text{abs.} \}, \end{aligned} \tag{7}$$

where abs. refers to absorption terms and $\mathcal{W}_g^{\text{em}}(\mathbf{Q}) = |\mathcal{M}_g(\mathbf{Q})|^2 (N_g + 1)$, with E_g the phonon energy and $N_g = (e^{E_g/k_B T} - 1)^{-1}$ the phonon occupation. The absorption terms can be obtained in the end by flipping the sign of the phonon energy E_g and making the switch $N_g + 1 \rightarrow N_g$. The two terms in Eq. (7) correspond to out-scattering (first term, negative sign) and in-scattering (second term, positive sign). The appendix gives $\mathcal{W}_g^{\text{em}}(\mathbf{Q})$ for various interaction mechanisms.

To proceed further, we pick the eigenstates of \hat{H}_0 as a basis. The eigenstates are denoted as $|n, \mathbf{k}\rangle = |n\rangle \otimes |\mathbf{k}\rangle$, where n labels the discrete set of eigenfunctions with energy E_n in the z -direction (subband energies) and \mathbf{k} labels the continuous set of free-particle eigenfunctions with energy $E_k = \hbar^2 k^2 / 2m^*$ in the in-plane direction with the effective mass m^* . The phase of the basis states is chosen such that $\langle z|n\rangle$ are real. With this choice of basis, we have $\langle n', \mathbf{k}'|n, \mathbf{k}\rangle = \delta_{n'n} \delta(\mathbf{k}' - \mathbf{k})$. We assume translational invariance in the in-plane direction so both the density matrix and the dissipator are diagonal in \mathbf{k}

$$\langle n', \mathbf{k}'|\hat{\rho}_e|n, \mathbf{k}\rangle = \rho_{n'n}^{E_k} \delta(\mathbf{k}' - \mathbf{k}), \tag{8a}$$

$$\langle n', \mathbf{k}'|\hat{D}|n, \mathbf{k}\rangle = \mathcal{D}_{n'n}^{E_k} \delta(\mathbf{k}' - \mathbf{k}), \tag{8b}$$

where the matrix elements of $\hat{\rho}_e$ and \hat{D} are labeled according to their in-plane kinetic energy E_k . In order to make the following derivation more compact, we define the following quantities:

$$(n|m)_{Q_z} = \langle n|e^{iQ_z \hat{z}}|m\rangle, \tag{9a}$$

$$\Delta_{nm} = E_n - E_m, \tag{9b}$$

$$E(n, \mathbf{k}) = E_n + E_k. \tag{9c}$$

In Sect. 2.1, we simplify the out-scattering term in Eq. (7) and do the same for the in-scattering term in Sect. 2.2. In Sect. 2.3, we write the master equation in a form applicable to periodic systems such a QCLS.

2.1 Out-scattering term

We start with the out-scattering-term, which is the first term in Eq. (7). By using the completeness relation 4 times, we can write the dissipator term corresponding to emission due to interaction mechanism g as

$$\begin{aligned} \hat{D}_{\text{em},g}^{\text{out}} = & -\frac{1}{\hbar^2 (2\pi)^3} \int d^3 Q \int_0^\infty ds \sum_{n_{1234}} \int d^2 k_{1234} \\ & \times \langle n_1, \mathbf{k}_1|e^{-i\mathbf{Q}\cdot\hat{\mathbf{R}}} e^{-i\hat{H}_0 s/\hbar}|n_2, \mathbf{k}_2\rangle e^{-iE_g s/\hbar} \\ & \times \langle n_2, \mathbf{k}_2|e^{+i\mathbf{Q}\cdot\hat{\mathbf{R}}} e^{+i\hat{H}_0 s/\hbar}|n_3, \mathbf{k}_3\rangle \mathcal{W}_g^{\text{em}}(\mathbf{Q}) \\ & \times \langle n_3, \mathbf{k}_3|\hat{\rho}_e|n_4, \mathbf{k}_4\rangle \langle n_1, \mathbf{k}_1|n_4, \mathbf{k}_4\rangle + \text{h.c.}, \end{aligned} \tag{10}$$

where $\int d^2 k_{1234}$ refers to integration over \mathbf{k}_1 through \mathbf{k}_4 and n_{1234} refers to sum over n_1 through n_4 . We can simplify the above expression using

$$\begin{aligned} & \langle n, \mathbf{k}|e^{\pm i\mathbf{Q}\cdot\hat{\mathbf{R}}} e^{\pm i\hat{H}_0 s/\hbar}|n', \mathbf{k}'\rangle \\ & = (n|n')_{Q_z}^* e^{\pm iE(n', \mathbf{k}')s/\hbar} \delta[\mathbf{k} - (\mathbf{k}' \pm \mathbf{q})]. \end{aligned} \tag{11}$$

Using Eq. (11) and after performing the \mathbf{k}_4 integration, Eq. (10) becomes

$$\begin{aligned} \hat{D}_{\text{em},g}^{\text{out}} = & -\frac{1}{\hbar^2 (2\pi)^3} \int \int d^3 Q \int_0^\infty ds \sum_{n_{1234}} d^2 k_{123} \\ & \times \mathcal{W}_g^{\text{em}}(\mathbf{Q}) (n_1|n_2)_{Q_z}^* (n_2|n_3)_{Q_z} \\ & \times e^{-i\frac{s}{\hbar}(E(n_2, \mathbf{k}_2) - E(n_3, \mathbf{k}_3) + E_g)} \rho_{n_3 n_4}^{E_{k_3}} |n_1, \mathbf{k}_1\rangle \langle n_4, \mathbf{k}_3| \\ & \times \delta[\mathbf{k}_1 - (\mathbf{k}_2 - \mathbf{q})] \delta[\mathbf{k}_2 - (\mathbf{k}_3 + \mathbf{q})] + \text{h.c.}. \end{aligned} \tag{12}$$

After performing the \mathbf{k}_2 and \mathbf{k}_3 integration, we get

$$\begin{aligned} \hat{D}_{\text{em},g}^{\text{out}} = & -\frac{1}{\hbar^2 (2\pi)^3} \int d^3 Q \int_0^\infty ds \sum_{n_{1234}} \int d^2 k_1 \\ & \times \mathcal{W}_g^{\text{em}}(\mathbf{Q}) (n_1|n_2)_{Q_z}^* (n_2|n_3)_{Q_z} \rho_{n_3 n_4}^{E_{k_1}} |n_1, \mathbf{k}_1\rangle \langle n_4, \mathbf{k}_1| \\ & \times e^{-i\frac{s}{\hbar}(E(n_2, \mathbf{k}_1 + \mathbf{q}) - E(n_3, \mathbf{k}_1) + E_g)} + \text{h.c.}. \end{aligned} \tag{13}$$

In order to perform the s integration, we use

$$\int_0^\infty e^{-i\Delta \frac{s}{\hbar}} ds = \pi \hbar \delta(\Delta) - i \hbar \mathcal{P} \frac{1}{\Delta}, \tag{14}$$

where \mathcal{P} denotes the Cauchy principal value, which leads to a small correction to energies (Lamb shift) [18]. Ignoring the principal-value term and shifting the integration variable $\mathbf{Q} \rightarrow \mathbf{Q} - \mathbf{k}_1$, we get

$$\begin{aligned} \hat{D}_{em,g}^{\text{out}} = & -\frac{\pi}{\hbar(2\pi)^3} \int d^3 Q \sum_{n_{1234}} \int d^2 k_1 \\ & \times \mathcal{W}_g^{\text{em}}(\mathbf{Q} - \mathbf{k}_1) (n_1|n_2)_{Q_z}^* (n_2|n_3)_{Q_z} \rho_{n_3 n_4}^{E_{k_1}} \\ & \times \delta[\Delta_{n_2 n_3} + E_q - E_{k_1} + E_g] |n_1, \mathbf{k}_1\rangle \langle n_4, \mathbf{k}_1| \\ & + \text{h.c.} \end{aligned} \tag{15}$$

Sandwiching both sides by $\langle N, \mathbf{k} | \dots | M, \mathbf{k}' \rangle$, integrating over \mathbf{k}_1 and \mathbf{k}' and renaming the sum variables $n_2 \rightarrow m, n_3 \rightarrow m$ gives

$$\begin{aligned} \left[\hat{D}_{em,g}^{\text{out}} \right]_{NM}^{E_k} = & - \sum_{n,m} \rho_{nM}^{E_k} \frac{\pi}{\hbar(2\pi)^3} \int d^3 Q \mathcal{W}_g^{\text{em}}(\mathbf{Q} - \mathbf{k}) \\ & \times \delta[\Delta_{mn} + E_q - E_k + E_g] (N|m)_{Q_z}^* (m|n)_{Q_z} \\ & + \text{h.c.}, \end{aligned} \tag{16}$$

where in this context, h.c. means “switch N and M and perform complex conjugation.” We can write Eq. (16) more compactly as

$$\left[\hat{D}_{em,g}^{\text{out}} \right]_{NM}^{E_k} = - \sum_n \rho_{nM}^{E_k} \Gamma_{em,g}^{\text{out}}(N, n, E_k) + \text{h.c.}, \tag{17}$$

with

$$\begin{aligned} \Gamma_{em,g}^{\text{out}}(N, n, E_k) = & \frac{\pi}{\hbar(2\pi)^3} \sum_m \int d^3 Q \mathcal{W}_g^{\text{em}}(\mathbf{Q}, E_k) \\ & \times \delta[\Delta_{mn} + E_q - E_k + E_g] (N|m)_{Q_z}^* (m|n)_{Q_z}, \end{aligned} \tag{18}$$

where we have written $\mathcal{W}_g^{\text{em}}(\mathbf{Q} - \mathbf{k}) = \mathcal{W}_g^{\text{em}}(\mathbf{Q}, E_k)$ because the coordinate system for the \mathbf{q} integration can be chosen relative to \mathbf{k} so $\mathcal{W}_g^{\text{em}}$ only depends on the magnitude of \mathbf{k} . Note that $\Gamma_{em,g}^{\text{out}}$ has the units of inverse time and is real. These terms will be referred to as *rates* from now on. The rates do not depend on the density matrix, so they can be precalculated and stored.

The Q_z integration in Eq. (18) involves inner products, such as $(m|n)_{Q_z}$, and has to be performed numerically. However, the in-plane integration can be done analytically, so it is useful to rewrite Eq. (18) as

$$\Gamma_{em,g}^{\text{out}}(N, n, E_k) = \frac{\pi}{\hbar(2\pi)^3} \sum_m \int_{-\infty}^\infty dQ_z (N|m)_{Q_z}^*$$

$$\times (m|n)_{Q_z} \int d^2 q \delta[\Delta_{nm} + E_k - E_g - E_q] \mathcal{W}_g^{\text{em}}(\mathbf{Q}, E_k). \tag{19}$$

The real (imaginary) part of the integrand is even (odd), so we can limit the range of integration to positive Q_z . Switching to polar coordinates $d^2 q \rightarrow q dq d\theta$, making a change of variables $E_q = \hbar^2 q^2 / 2m^*$ and performing the E_q integration gives

$$\begin{aligned} \Gamma_{em,g}^{\text{out}}(N, n, E_k) = & \frac{m^*}{2\pi \hbar^3} \sum_m \theta(\Delta_{nm} - E_g + E_k) \\ & \times \int_0^\infty dQ_z \text{Re} \left[(N|m)_{Q_z}^* (m|n)_{Q_z} \right] \\ & \times \mathcal{G}_g^{\text{em}}(E_k, Q_z, \Delta_{nm} - E_g + E_k), \end{aligned} \tag{20}$$

where θ is the Heaviside function and

$$\mathcal{G}_g^{\text{em}}(E_k, Q_z, E_q) = \frac{1}{2\pi} \int_0^{2\pi} d\theta \mathcal{W}_g^{\text{em}}(\mathbf{Q}, E_k). \tag{21}$$

$\mathcal{W}(\mathbf{Q}, E_k)$ can always be written in terms of $Q_z, E_k, E_q = \hbar^2 q^2 / 2m^*$ and the polar angle θ of \mathbf{q} . The explicit form of the function $\mathcal{G}_g^{\text{em}}$ depends on the scattering mechanism g , and is calculated in Appendix for acoustic phonons, nonpolar optical phonons, polar optical phonons (POP), and ionized impurities.

2.2 In-scattering term

By using the completeness relation four times, the in-scattering term in Eq. (7) becomes

$$\begin{aligned} \hat{D}_{em,g}^{\text{in}} = & \frac{1}{\hbar^2(2\pi)^3} \int d^3 Q \int_0^\infty ds \sum_{n_{1234}} \int d^2 k_{1234} \\ & \times \mathcal{W}_g^{\text{em}}(\mathbf{Q}) e^{iE_g s/\hbar} \langle n_1, \mathbf{k}_1 | e^{i\mathbf{Q}\cdot\hat{\mathbf{r}}} | n_2, \mathbf{k}_2 \rangle \\ & \times \langle n_3, \mathbf{k}_3 | e^{-i\hat{H}_0 s/\hbar} e^{-i\mathbf{Q}\cdot\hat{\mathbf{r}}} e^{i\hat{H}_0 s/\hbar} | n_4, \mathbf{k}_4 \rangle \\ & \times \langle n_2, \mathbf{k}_2 | \hat{\rho}_e | n_3, \mathbf{k}_3 \rangle | n_1, \mathbf{k}_1 \rangle \langle n_4, \mathbf{k}_4 | + \text{h.c.} \end{aligned} \tag{22}$$

Using Eq. (8a) and

$$\begin{aligned} & \langle n_3, \mathbf{k}_3 | e^{-i\hat{H}_0 s/\hbar} e^{-i\mathbf{Q}\cdot\hat{\mathbf{r}}} e^{i\hat{H}_0 s/\hbar} | n_4, \mathbf{k}_4 \rangle \\ & = e^{-i\frac{s}{\hbar}(E(n_3, \mathbf{k}_3) - E(n_4, \mathbf{k}_4))} (n_3|n_4)_{Q_z}^* \delta[\mathbf{k}_4 - (\mathbf{k}_3 + \mathbf{q})], \end{aligned} \tag{23}$$

gives (after performing the \mathbf{k}_4 integration)

$$\begin{aligned} \hat{D}_{em,g}^{\text{in}} = & \frac{1}{\hbar^2(2\pi)^3} \int d^3 Q \int_0^\infty ds \sum_{n_{1234}} \int d^2 k_{123} \\ & \times \mathcal{W}_g^{\text{em}}(\mathbf{Q}) (n_1|n_2)_{Q_z} (n_3|n_4)_{Q_z}^* \end{aligned}$$

$$\begin{aligned} &\times \rho_{n_2 n_3}^{E_{k_2}} e^{-i \frac{s}{\hbar} (E(n_3, \mathbf{k}_3) - E(n_4, \mathbf{k}_3 + \mathbf{q}) - E_g)} \delta[\mathbf{k}_2 - \mathbf{k}_3] \\ &\times |n_1, \mathbf{k}_1\rangle \langle n_4, \mathbf{k}_3 + \mathbf{q}| + \text{h.c.} \end{aligned} \tag{24}$$

Performing the \mathbf{k}_3 and \mathbf{k}_2 integrations gives

$$\begin{aligned} \hat{\mathcal{D}}_{\text{em},g}^{\text{in}} &= \frac{1}{\hbar^2 (2\pi)^3} \int d^3 Q \int_0^\infty ds \sum_{n_{1234}} \\ &\times \int d^2 k_1 \mathcal{W}_g^{\text{em}}(\mathbf{Q})(n_1|n_2)_{Q_z} (n_3|n_4)_{Q_z}^* \rho_{n_2 n_3}^{E_{|\mathbf{k}_1 - \mathbf{q}|}} \\ &\times e^{-i \frac{s}{\hbar} (E(n_3, \mathbf{k}_1 - \mathbf{q}) - E(n_4, \mathbf{k}_1) - E_g)} |n_1, \mathbf{k}_1\rangle \langle n_4, \mathbf{k}_1| \\ &+ \text{h.c.} \end{aligned} \tag{25}$$

Changing the \mathbf{Q} integration variable $\mathbf{Q} \rightarrow -\mathbf{Q} + \mathbf{k}_1$ and performing the s integration (ignoring the principal value) gives

$$\begin{aligned} \hat{\mathcal{D}}_{\text{em},g}^{\text{in}} &= \frac{\pi}{\hbar (2\pi)^3} \int d^3 Q \sum_{n_{1234}} \int d^2 k_1 \mathcal{W}_g^{\text{em}}(\mathbf{Q} - \mathbf{k}_1) \\ &\times (n_1|n_2)_{Q_z} (n_3|n_4)_{Q_z}^* \rho_{n_2 n_3}^{E_q} \delta[\Delta_{n_3 n_4} - E_{k_1} - E_g + E_q] \\ &\times |n_1, \mathbf{k}_1\rangle \langle n_4, \mathbf{k}_1| + \text{h.c.}, \end{aligned} \tag{26}$$

where we have used $\mathcal{W}_g^{\text{em}}(-\mathbf{Q} + \mathbf{k}_1) = \mathcal{W}_g^{\text{em}}(\mathbf{Q} - \mathbf{k}_1)$. Sandwiching both sides by $\langle N, \mathbf{k} | \dots | M, \mathbf{k}' \rangle$, integrating over \mathbf{k}' and \mathbf{k}_1 and renaming the dummy variables $n_2 \rightarrow n$ and $n_3 \rightarrow m$ gives

$$\begin{aligned} \left[\hat{\mathcal{D}}_{\text{em},g}^{\text{in}} \right]_{NM}^{E_k} &= \sum_{n,m} \rho_{nm}^{E_k + E_g + \Delta_{Mm}} \frac{\pi}{\hbar (2\pi)^3} \int d^3 Q \\ &\times \mathcal{W}_g^{\text{em}}(\mathbf{Q} - \mathbf{k})(N|n)_{Q_z} (m|M)_{Q_z}^* \\ &\times \delta[\Delta_{Mm} + E_k + E_g - E_q] + \text{h.c.} \end{aligned} \tag{27}$$

After doing the in-plane integration over \mathbf{q} , we get

$$\begin{aligned} \left[\hat{\mathcal{D}}_{\text{em},g}^{\text{in}} \right]_{NM}^{E_k} &= \sum_{n,m} \rho_{nm}^{E_k + E_g + \Delta_{Mm}} \Gamma_{\text{em},g}^{\text{in}}(N, M, n, m, E_k) \\ &+ \text{h.c.}, \end{aligned} \tag{28}$$

where we have defined the in-scattering analogue of Eq. (20)

$$\begin{aligned} \Gamma_{\text{em},g}^{\text{in}}(N, M, n, m, E_k) &= \frac{m^*}{2\pi \hbar^3} \theta[\Delta_{Mm} + E_g + E_k] \\ &\times \int_0^\infty dQ_z \text{Re} \left[(N|n)_{Q_z} (m|M)_{Q_z}^* \right] \\ &\times \mathcal{G}_g^{\text{em}}(E_k, Q_z, \Delta_{Mm} + E_g + E_k), \end{aligned} \tag{29}$$

with $\mathcal{G}_g^{\text{em}}$ defined in Eq. (21).

2.3 Application to periodic systems

The Markovian master equation (MME) for the density-matrix elements can be written by summing over all different scattering mechanisms g :

$$\frac{\partial \rho_{NM}^{E_k}}{\partial t} = -i \frac{\Delta_{NM}}{\hbar} \rho_{NM}^{E_k} + \mathcal{D}_{NM}^{E_k}, \tag{30a}$$

with

$$\begin{aligned} \mathcal{D}_{NM}^{E_k} &= - \sum_{g,n} \Gamma_{\text{em},g}^{\text{out}}(N, n, E_k) \rho_{nM}^{E_k} \\ &+ \sum_{n,m,g} \Gamma_{\text{em},g}^{\text{in}}(N, M, n, m, E_k) \rho_{nm}^{E_k + E_g + \Delta_{Mm}} \\ &+ \text{h.c.} + \text{abs.}, \end{aligned} \tag{30b}$$

where abs. refers to absorption terms and the in- and out-scattering rates are defined in Eqs. (20) and (29), respectively. The $\mathcal{G}_g^{\text{em}}$ functions are calculated in Appendix for various scattering mechanisms.

The form of the MME in Eqs. (30) is not well suited for periodic systems such as QCLs. It is more convenient to work with relative indices:

$$f_{N,M}^{E_k} \equiv \rho_{N,N+M}^{E_k}. \tag{31}$$

Using relative indices, it is easy to take advantage of periodicity, where

$$f_{N,M}^{E_k} = f_{N \pm N_s, M}^{E_k}. \tag{32}$$

The range $N \in [1, N_s]$ is the number of eigenstates N_s in a single stage. The choice of which stage to consider is arbitrary but in this work we choose the central stage corresponding to the range $z \in [-L_p/2, L_p/2]$, where L_p is the stage length. A state is considered to be in the central stage if $|\langle n | \hat{z} | n \rangle| \leq L_p/2$, i.e., if the state's expectation value of cross-plane position, $\langle z \rangle_n$, is within the central stage. The elements with $M = 0$ give the diagonals of the density matrix and $M \neq 0$ gives the coherence a distance of M from the diagonal. The M index runs from $-\infty$ to $+\infty$ so a truncation needs to be performed in order to do numerical calculations. Truncation of M will be discussed in Sect. 3.1.

Inserting Eq. (31) into Eq. (30) gives

$$\begin{aligned} \frac{\partial f_{N,M}^{E_k}}{\partial t} &= -i \frac{\Delta_{N,N+M}}{\hbar} f_{N,M}^{E_k} - \sum_{n,g} \Gamma_{NMnE_k}^{\text{out,em},g} f_{N,n}^{E_k} \\ &+ \sum_{n,m,g} \Gamma_{NMnmE_k}^{\text{in,em},g} f_{N+n,M+m-n}^{E_k + E_g + \Delta_{N+Mn+M+m}} + \text{h.c.} + \text{abs.} \end{aligned} \tag{33}$$

with

$$\Gamma_{NMnE_k}^{\text{out,em}} = \frac{m^*}{2\pi\hbar^3} \sum_m \theta[\Delta_{N+n,N+M+m} - E_g + E_k] \int_0^\infty \times dQ_z \text{Re} \left[(N+M|N+M+m)_{Q_z}^* (N+M+m|N+n)_{Q_z} \right] \times \mathcal{G}_g^{\text{em}}(E_k, Q_z, \Delta_{N+n,N+M+m} - E_g + E_k), \tag{34a}$$

and

$$\Gamma_{NMnmE_k}^{\text{in,em,g}} = \frac{m^*}{2\pi\hbar^3} \theta[\Delta_{N+M,N+M+m} + E_g + E_k] \times \int_0^\infty dQ_z \text{Re} \left[(N|N+n)_{Q_z} (N+M+m|N+M)_{Q_z}^* \right] \times \mathcal{G}_g^{\text{em}}(E_k, Q_z, \Delta_{N+M,N+M+m} + E_g + E_k). \tag{34b}$$

Note that, in Eqs. (34a) and (34b), the dummy indices n and m have been shifted in such a way that terms with large n or m are small. Equation (33) [with accompanying Eqs. (34a) and (34b)] is the main result in this work. In the next section, we will discuss numerical solution methods for Eq. (33). For evaluation of $\mathcal{G}_g^{\text{em}}$ for various interaction mechanisms, we refer the reader to the Appendix.

Equation (33) improves upon previous theoretical work on QCLs [5,11,15], where transport was considered as semiclassical within a single period (or a subregion of a single period), while coupling between different stages (separated by a thick injector barrier) was treated quantum mechanically, using phenomenological dephasing times. Here, we derived a completely positive Markovian evolution for the density matrix, which supplants the need for phenomenological dephasing times and captures all relevant coherent transport features, not just tunneling between stages. We note that Eq. (33) can be derived as the low-density limit of the more general but still completely positive Markovian equation that includes electron–electron interaction [22]; in THz QCLs, carrier density is low, so the low-density limit is appropriate. We also note that our main equation (33) (with (34a) and (34b)) is equivalent to the low-density, Markovian, single-particle density-matrix equations Eqs. (112)–(116) from Iotti et al. [28]; however, Iotti et al. derived these from a more general Markovian set of equations that are not completely positive [28]. We also emphasize that the formulation presented here is tailored toward efficient numerical implementation.

3 Numerical method

The central quantity is the density matrix $f_{NM}^{E_k}$, which is stored for $N \in [1, N_s]$, $M \in [-N_c, N_c]$, $E_k \in [0, E_{\text{max}}]$. Here, N_c is an integer that quantifies how far apart in energy the states can be to still have appreciable off-diagonal

density-matrix terms (coherences); we refer to N_c as the coherence cutoff. E_{max} is the in-plane kinetic-energy cutoff. The energies are discretized into N_E evenly spaced values, such that the density-matrix array has dimensions $N_s \times (2N_c + 1) \times N_E$.

The basic idea is to start with a chosen initial state and numerically time-step Eq. (33), until a steady state is reached. For the time stepping, we use an asynchronous leapfrog method, which is a robust second order, two-step, explicit method for the integration of the Liouville equation [23]. This choice of the time-stepping method allows us to use a rather large time step of 1 fs, which is about 10 times larger than an Euler time-stepping scheme would allow.

Note that the sums in the MME (33) run over the matrix elements and energies outside the fundamental period (e.g., $N > N_s$ or $N < 1$), which are calculated using the modulo operation:

$$f_{N,M}^{E_k} = f_{N',M}^{E_k} \tag{35a}$$

$$E_N = E_{N'} + \frac{N - N'}{N_s} E_0 \tag{35b}$$

$$N' = \text{mod}(N - 1, N_s) + 1, \tag{35c}$$

with $\text{mod}(n, N_s) = n - N_s \lfloor n/N_s \rfloor$ and E_0 the potential energy drop over a single period (intrinsic function MOD in Matlab and MODULUS in gfortran). The MME (33) also contains terms for which $|M| > N_c$, where we assume $f_{NM}^{E_k} = 0$.

3.1 Coherence cutoff and performance

Equation (33) contains an infinite sum that represents coupling between eigenstates over infinitely long distances. However, it is easy to see that terms with small $|n|$ and $|m|$ are dominant. For example, the in-scattering term contains terms on the form $(N|N+n)_{Q_z}$ and $(N+M+m|N+M)_{Q_z}^*$, which are small for large $|n|$ and $|m|$, respectively, due to the low spatial overlap of the states that are highly separated in energy. For the same reason, out-scattering terms with high $|n|$ or $|m|$ are small, too. In this work, we truncate the sum by only including terms with $|n|, |m| \leq N_c$. We note that the numerical method could be improved by only summing over a subset of $n, m \in [-N_c, N_c]$, that contains the biggest rates.

From Eq. (33), we see that the in-scattering term is the bottleneck in the time-evolution of the density matrix. The computational complexity for the time-evolution is $\mathcal{O}(N_s N_c^3 N_E N_g)$, and therefore depends most strongly on the coherence cutoff, N_c . The computational complexity only depends linearly on the number of eigenstates N_s , which opens the possibility to study multiple periods of QCLs and investigate effects of electric field domain formation [24], which has a negative effect of QCL performance.

The minimum coherence cutoff needed for convergence is highly system dependent. In this work, a modest value of $N_c = 5$ proved to be sufficient for convergence in current and occupations. Other parameters used in this work are $N_E = 101$, $N_s = 5$, and $N_g = 4$. The number of time steps is 10^5 , with a time step of 1 fs, resulting in 100 ps of simulated time. Using these parameters, typical simulation times for a single value of the electric field were about 45 minutes on an Intel Core i7-2600 (gfortran compiler, running on a single core). As mentioned before, the simulation time could be reduced significantly by only summing over a chosen small subset of $n, m \in [-N_c, N_c]$ in Eq. (33).

3.2 Initial state

We choose an initial state corresponding to thermal equilibrium. Assuming Boltzmann statistic, the density-matrix factors into the in-plane and cross-plane terms and we can write

$$f_{NM}^{E_k} \Big|_{\text{eq}} = C_{NM} e^{-E_k/k_B T}. \tag{36}$$

To calculate the expansion coefficients C_{NM} , we first solve for the Bloch states $\phi_{s,q}(z)$ (s labels the band and $q \in [-\pi/L_p, \pi/L_p]$ labels the wave vector in the Brillouin zone associated with the structure’s period L_p) by diagonalizing the Hamiltonian in (39) with $V_B(z) = 0$, using a basis of plane waves. We can then calculate the cross-plane equilibrium density matrix using

$$\rho_{\text{eq}}(z_1, z_2) = \sum_s \int_{-\pi/L_p}^{\pi/L_p} \phi_{s,q}(z_1) \phi_{s,q}^*(z_2) e^{-E_{s,q}/k_B T} dq. \tag{37}$$

Using the above result, we can calculate the expansion coefficients

$$C_{NM} = \int dz_1 dz_2 \psi_N(z_1) \psi_{N+M}(z_2) \rho_{\text{eq}}(z_1, z_2). \tag{38}$$

This choice of initial condition works well with an electric field that is turned on instantaneously at time $t = 0^+$; this is the limiting case of an abruptly turned-on bias. If only the steady state is sought, all terms with $M \neq 0$ can be artificially set equal to zero in the initial density matrix; this initial condition avoids high-amplitude coherent oscillations during the transient and leads to a faster numerical convergence toward the steady state.

3.3 Bandstructure calculation

Upon the application of bias, we assume the field and the associated linear potential drop are established instantaneously,

but that the density matrix and charge distribution take a while to respond and do so adiabatically.

We treat the eigenstates under an applied bias as bound states, even though, strictly speaking, the states are better described as resonances with some energy spread [25]. The bound-state approximation is good if the energy spread is much smaller than other characteristic energies, and if the dynamics are mostly limited to the subspace of resonance states. For more discussion on the validity of this approximation, see Ref. [3].

The eigenstates and subband energies are obtained from the Schrödinger equation

$$\left(-\frac{\hbar^2}{2} \frac{d}{dz} \frac{1}{m(z)} \frac{d}{dz} + V_{\text{SL}}(z) + V_{\text{B}}(z) + V_{\text{H}}(z, t) \right) \psi_n(z, t) = E_n(t) \psi_n(z, t), \tag{39}$$

where $m^*(z)$ is a position-dependent effective mass, V_{SL} is the superlattice potential (wells and barriers), V_{B} the linear potential drop due to an applied bias, and V_{H} the mean-field Hartree potential, which is obtained by solving Poisson’s equation. The Hartree potential V_{H} depends on the electron density and is therefore time-dependent. However, its time evolution is weak due to low doping and is typically very slow, so we can assume that the adiabatic approximation holds and the concept of eigenstates and energies is well defined during the transient. In writing Eq. (39), we have neglected coupling of the eigenfunctions with the in-plane motion, which is a standard assumption when describing QCLs and superlattices [3, 24].

To calculate the eigenfunctions under bias, which we assume are fairly well localized, we use a basis of Hermite functions (eigenfunctions of the harmonic oscillator) and diagonalize the Hamiltonian in Eq. (39). We solve the Schrödinger equation over a simulation domain that contains nine stages of the considered QCL in order to obtain reliable eigenstates in the central stage and minimize the effects due to the proximity of the boundaries. States in other periods are obtained via translation of the central-stage states and the number of relevant stages for the construction of the density matrix depends on the choice of the coherence cutoff N_c . This bound-state approach is known to produce spurious solutions, unphysical states that have high amplitude far above the conduction band edge, which we systematically eliminate. First, we discard the states that have energies higher than $E_{\text{max}} = 0.5$ eV above the conduction band top or lower than $E_{\text{min}} = 0.3$ eV below the conduction band bottom in the central stage; this restriction reduces the number of states from about 600 (number of basis states used) to a more manageable number of about 50. Second, for a state with index n among these remaining states, we then calculate two quantities, $z_{L,n}$ and $z_{R,n}$, defined as $\int_{z_{L,n}}^{\infty} dz |\Psi_n(z)|^2 = 99.9\%$ and

$\int_{z_{R,n}}^{-\infty} dz |\Psi_n(z)|^2 = 99.9\%$. The $\mp\infty$ refer to the left/right boundary of the nine-stage simulation domain. We discard the states that have a spread $z_{R,n} - z_{L,n}$ greater than a certain threshold value L_S or have more than N_{nodes} between $z_{L,n}$ and $z_{R,n}$. In this work, we use $L_S = 100$ nm, which is between two and three stage lengths, and $N_{\text{nodes}} = 20$. This method removes fast-oscillating and delocalized states. Finally, to prevent duplicates of the same eigenstate corresponding to different periods, we calculate the expectation value of z for each state and discard the states for which it is outside of the central period. For the remaining states, we follow [3] and sort the states in order of increasing average energy with respect to the conduction band edge, then keep the first N_s . N_s must be high enough to include all states that are localized below the top of the potential barriers. (Including too many states above the conduction band has a minimal effect on transport, since these states are mostly unoccupied, but needlessly increases computational overhead.)

As the system evolves, the eigenfunctions and rates (Eqs. (34a) and (34b)) have to be recalculated, which is computationally expensive, taking about 10 to 50 times longer than a single time step. However, it does not need to be done in every time step due to the slow temporal and spatial variation of V_H . In order to recalculate the eigenfunctions only when needed, we calculate

$$\delta = \max_m \left| \int dz |\psi_m(z, t_\ell)|^2 (V_H(z, t_\ell) - V_H(z, t_{\ell-1})) \right|, \tag{40}$$

where t_ℓ is the time at the current time step ℓ . The quantity δ is the magnitude of the maximal first-order energy correction to the eigenstates. If δ is above a certain threshold energy, we recalculate the eigenfunctions and the corresponding rates. If the threshold is not met, we do not update the wavefunctions nor the Hartree potential. The procedure of calculating δ is very cheap in terms of computational resources and does not noticeably affect performance. Typically, the wavefunctions are recalculated frequently during the initial transient and much less frequently near the steady state. In the present work, we used a threshold energy of 0.1 meV. This choice of threshold energy typically leads to ~ 100 recalculations of eigenfunctions while the total number of time steps is on the order of 10^5 .

When recalculating the eigenfunctions, an issue arises when numbering the updated states and choosing their phase. The time evolution of the eigenfunctions must be adiabatic so the same phase must be chosen for each state when the eigenfunctions are recalculated. Since the eigenfunctions are chosen to be real, there are only two choices of phase. A very simple assigning method is to calculate

$$\alpha_{nm} = \int dz \psi_n(z, t_\ell) \psi_m(z, t_{\ell-1}). \tag{41}$$

States with the highest overlap $|\alpha_{nm}| \simeq 1$ are “matched” according to $n \rightarrow m$, which ensures the proper numbering of the new states and $\psi_n(z, t_\ell) \rightarrow \text{sign}(\alpha_{nm})\psi_n(z, t_\ell)$ takes care of the choice of phase.

3.4 Low-energy thermalization

Out of the included scattering mechanisms (POP, acoustic phonons, and ionized impurities) only POP scattering is inelastic. However, the POP energy is typically larger than $k_B T$ and this lack of a low-energy inelastic scattering mechanism leads to numerical difficulties, where in-plane energy distributions can vary abruptly (this problem is often encountered in density-matrix models; see, for example, Ref. [10]). A detailed inclusion of electron–electron interaction would solve this issue, where arbitrarily low energy can be exchanged between electrons. The small energy exchanges involved in electron–electron interaction also plays a crucial role in thermalization within a subband. However, electron–electron interaction is a two-body interaction that is not straightforward to include in a single-electron picture. For this reason, in the present work, we will include low-energy thermalization (LET) in a simplified manner, by adding a scattering mechanism with an energy equal to the minimal in-plane energy spacing ΔE in the simulation. The purpose of this extra scattering mechanism is to help smoothen the in-plane energy distribution. We treated the LET as an additional POP-like scattering term, with energy exchange equal to ΔE and an effective strength denoted by the dimensionless quantity α . The matrix element is

$$|\mathcal{M}_{\text{LET}}(\mathbf{Q})|^2 = \alpha \frac{e^2 \Delta E}{2\epsilon_0} \left(\frac{1}{\epsilon_r^\infty} - \frac{1}{\epsilon_r} \right) \frac{Q^2}{(Q^2 + Q_D^2)^2}, \tag{42}$$

where $Q_D^2 = ne^2/(\epsilon k_B T)$ is the Debye wave vector. The reason for choosing a POP-like matrix element is its preference for small- \mathbf{Q} scattering, just like electron–electron interaction. The role of the LET term is mainly to smoothen the in-plane energy distribution. As we will show later, the results are not very sensitive to the value of α .

4 Results

To demonstrate the validity of our model, we simulated a THz QCL proposed in Ref. [14]. The authors used a phonon-assisted injection and extraction design based on a GaAs/Al_{0.25}Ga_{0.75}As material system and achieved lasing at 3.2 THz, up to a heatsink temperature of 138 K. We chose this specific device because both experimental and theoretical results are readily available for comparison. Figure 1 shows the conduction band profile and most important eigenfunctions of the considered device at the design electric field of

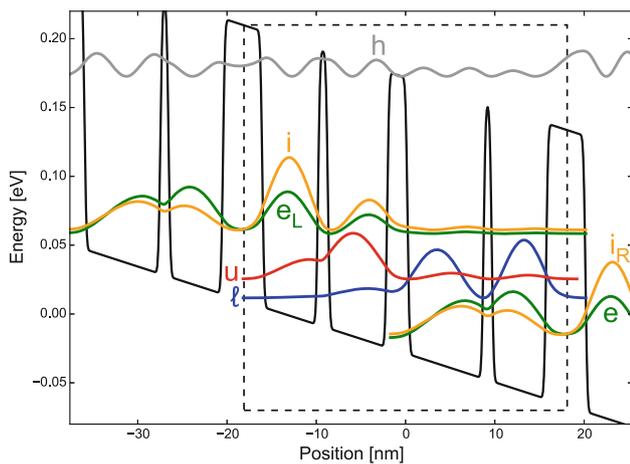


Fig. 1 Conduction band edge (*solid black line*) and probability densities for the upper lasing state (*u*), lower lasing state (*l*), injector state (*i*), and extractor state (*e*). Also shown is the extractor state (*e_L*) for the previous stage to the *left*, the injector (*i_R*) state for the next stage to the *right*, and a high-energy state (*h*). The high-energy state was included in numerical calculations; however, it had a small occupation and a negligible effect on physical observables. The *dashed rectangle* represents a single stage with the layer structure (from the *left*) **44/62.5/10.9/66.5/22.8/84.8/9.1/61** Å, with barriers in *bold font*. The *thickest barrier* (injector barrier) is doped with Si such that the average electron density is $8.98 \times 10^{15} \text{ cm}^{-3}$. Due to low doping, the potential drop is approximately uniform

21 kV/cm. We will split this section into two parts, starting with steady-state results in Sect. 4.1 and time resolved results in Sect. 4.2.

4.1 Steady-state results

Figure 2 shows a steady-state current density vs electric field, as well as comparison with experiment and theoretical results based on NEGF [14]. The experimental data is for a heat-sink temperature of $T_H = 10 \text{ K}$. The actual lattice temperature T_L is expected to be higher [7]. Both the density-matrix and NEGF results are for a lattice temperature of 50 K. We included interactions with polar optical phonons, acoustic phonons (using elastic and equipartition approximations), and ionized impurities. In addition, we included a LET scattering mechanism discussed in Sect. 3.4 with a strength parameter of $\alpha = 0.1$. This choice of α gave the best agreement with experiment. However, results around the design electric field did not depend strongly on α , as can be seen in Fig. 3. From Fig. 2 we see a very good agreement with experiment and NEGF around the design electric field of 21 kV/cm. For electric fields lower than 17 kV/cm, neither NEGF nor our density-matrix results accurately reproduce experimental results. However, our density-matrix results and the NEGF results both show a double-peak behavior. The difference between the density-matrix results and NEGF can be attributed to collisional broadening (not captured with

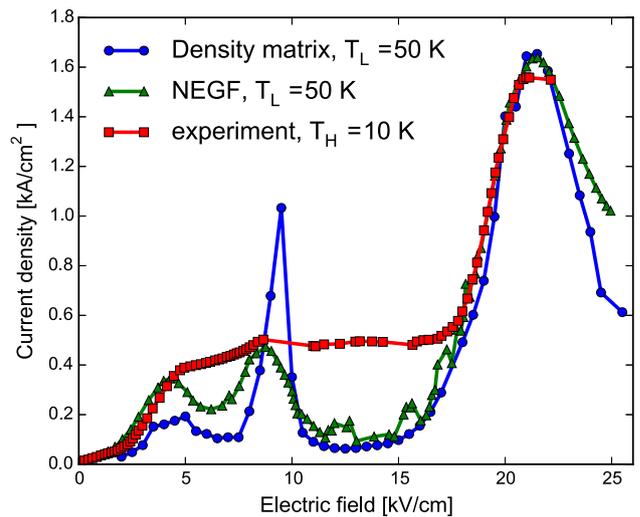


Fig. 2 Current density vs electric field for density matrix results (*blue circles*) and NEGF (*green triangles*) for a lattice temperature $T_L = 50 \text{ K}$. Also shown are experimental results (*red squares*) for a heat-sink temperature $T_H = 10 \text{ K}$. Experimental and NEGF results are both from Ref. [14] (Color figure online)

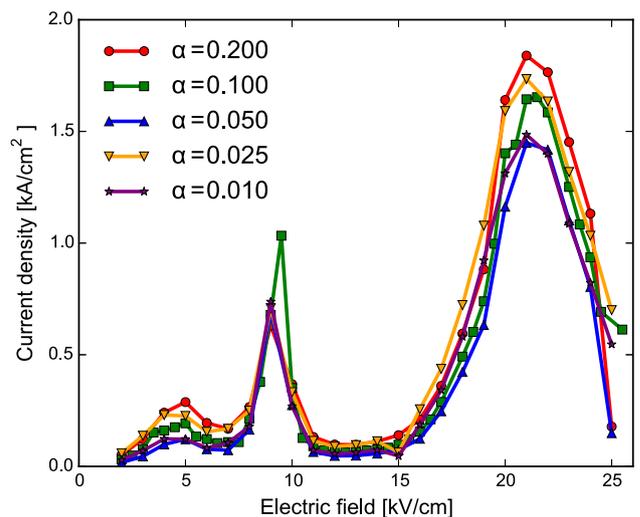


Fig. 3 Current density vs electric field for different values of the strength parameter α . The best agreement with experimental data is for $\alpha = 0.1$ (*green squares*). The results at high fields are not sensitive to the strength parameter, while low-bias results are. The higher peak at 9 kV/cm for the $\alpha = 0.1$ data is a result of a finer electric field mesh for that data set (Color figure online)

density matrix approaches) and our calculation not including interface-roughness scattering.

In order to visualize the occupations and coherences of all combinations of the states, it is instructive to plot density-matrix elements after integrating out the parallel energy

$$\rho_{NM} = \int \rho_{NM}^{E_k} dE_k. \tag{43}$$

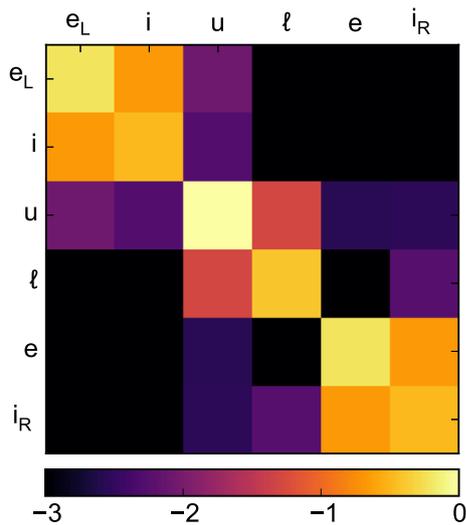


Fig. 4 $\log_{10}(|\rho_{NM}|)$, where ρ_{NM} are density-matrix elements after integration over parallel energy. Normalization is chosen so that the highest occupation is one (*upper lasing level*). Results are for an electric field of 21 kV/cm and a lattice temperature of 50 K. Coherences and occupations are given for all combinations of states shown in Fig. 1, except for the high energy h state, which had very small coherences and occupation. The highest occupations are the upper lasing level (1.0), extractor state (0.60), lower lasing level (0.39) and injector state (0.33). The biggest coherences are between the extractor state e and injector state i_R (0.21), and between the upper and lower lasing levels (0.05). Other coherences were smaller than 0.01

Figure 4 shows a plot of $\log_{10}(|\rho_{NM}|)$, with occupations and coherences of all combinations of the states shown in Fig 1, except for the high-energy (h) state, which had negligible occupation and coherences. Normalization is chosen such that the largest matrix element is 1 (the occupation of the upper lasing level). From the figure, we can see that the magnitude of the coherences can be quite large. For example, the largest coherence is between the e_L extractor state and the i injector state, with a magnitude of about 0.21; this is a significant fraction of the largest diagonal element and demonstrates the importance of including coherences in calculations. The second largest coherence is between the upper and lower lasing states, with a magnitude of 0.05. Other coherences are smaller than 1 % of the largest diagonal element and all coherences more than 4 places off the diagonal were smaller than 10^{-3} , justifying our coherence cutoff of $N_c = 5$.

The magnitude of the matrix elements ρ_{NM} gives information about the importance of including off-diagonal matrix elements in QCL simulations. However, these matrix elements do not give us information about the dependence on in-plane energy. In order to visualize the in-plane dependence, Fig. 5 shows plots of $\rho_{NM}^{E_k}$ as a function of the in-plane energy for multiple pairs of N and M . In the top (bottom) panel, $N = u (N = e)$ is fixed and M is varied; we show the three largest coherences $|\rho_{NM}^{E_k}|$, as well as the diagonal term

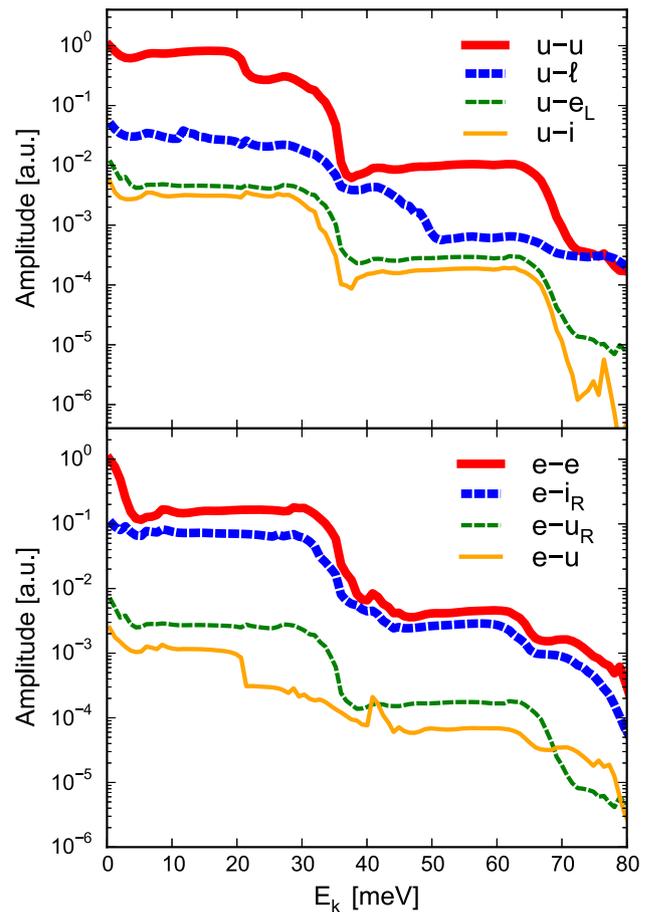


Fig. 5 The magnitude of the matrix element $|\rho_{NM}^{E_k}|$ as a function of in-plane energy, for N corresponding to the upper lasing level (*top panel*) and extractor state (*bottom panel*). These two states were chosen because they have the greatest occupations. Note that u_R corresponds to the upper lasing level in the next period to the right (not shown in Fig. 1), which is equal to the coherence between e_L and u , owing to periodicity

$|\rho_{NN}^{E_k}|$. The figure shows the energy dependence of the two largest coherences mentioned earlier ($e-i_R$ and $u-l$), along with the second and third largest coherences for each state. We see that most off-diagonal elements are more than two orders of magnitude smaller than the diagonal terms. Both the diagonal elements and the coherences have an in-plane distribution that deviates strongly from a Maxwellian distribution, with a sharp drop around 35 meV due to enhanced POP emission. This result suggests that simplified density-matrix approaches, where a Maxwellian in-plane distribution is assumed, are not justified for the considered system.

4.2 Time-resolved results

Figure 6 shows the current density vs time at the design electric field (21 kV/cm) and at a lower electric field (5 kV/cm). The top panel shows the initial transient (first 2 ps) and the

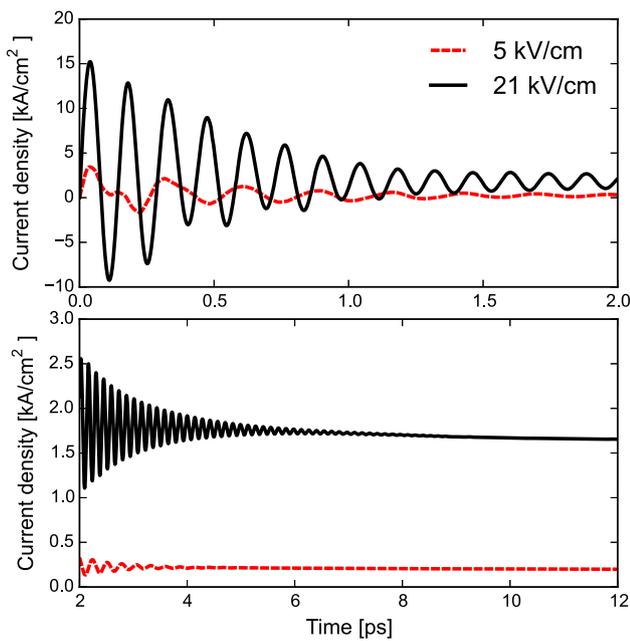


Fig. 6 Current density vs time for two values of electric fields. The upper panel shows the first 2 picoseconds and the lower the next 10 picoseconds. Note the different ranges on the vertical axis

bottom panel shows the next 10 ps, which is long enough for the current density to reach a steady state. In the first 2 ps, we observe high-amplitude coherent oscillations in current, with a period of 100–200 fs. The rapid coherent oscillations decay on a time scale of a few picoseconds, with the high-bias oscillations decaying more slowly. Note that the peak value of current early in the transient can be more than 10 times higher than the steady-state value. In the bottom panel, we see a slow change in current, which is related to the redistribution of electrons within subbands, as well as between different subbands.

Figure 7 shows the time evolution of occupations for the same values of bias as in Fig 6, in addition to results slightly below the design electric field. Occupations are very important quantities because the optical gain of the device is directly proportional to the population difference of the upper and lower lasing level ($\rho_{uu} - \rho_{\ell\ell}$). The time evolution of the occupations tells us how long it takes the device to reach its steady-state lasing capability. In Fig 6, we can see that the occupations take a much longer time to reach a steady state (20–100 ps) than the current density does, and the time needed to reach a steady state is not a monotonically increasing function of the electric field: the 20-kV/cm results take more than twice as long to reach a steady state than the 21-kV/cm results. In Fig. 7, we see that a population inversion of $\rho_{uu} - \rho_{\ell\ell} = 0.26$ is obtained at the design electric field, while lower-field results show no population inversion.

Figure 8 shows the time evolution of the in-plane energy distribution for all the subbands shown in Fig. 1. Also

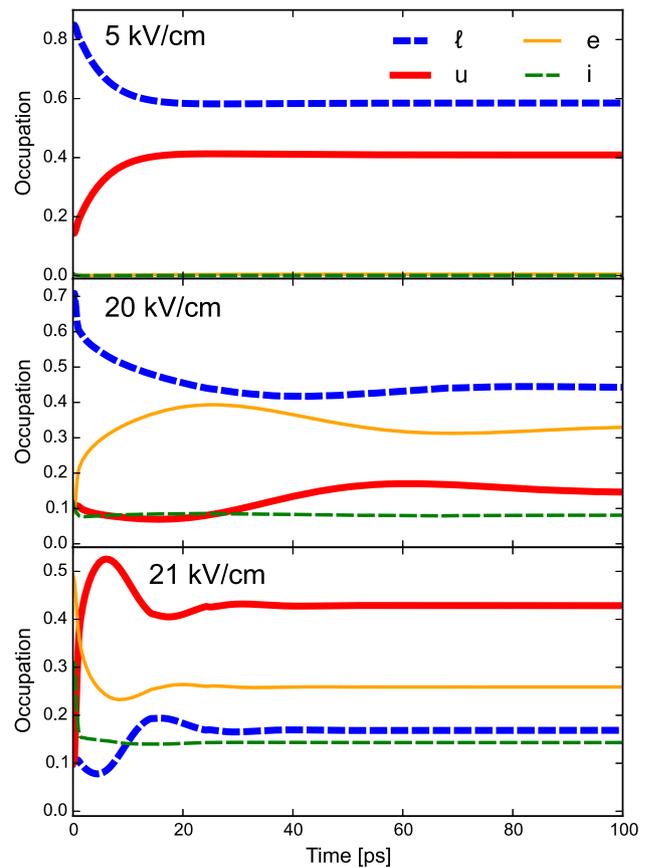


Fig. 7 Occupation vs time for the lower lasing, upper lasing, injector, and extractor states. Normalization is chosen such that all occupations add up to one. Note the longer time scale compared with the current density in Fig. 6

shown is the equivalent electron temperature of each subband calculated using $\langle E_k \rangle = k_B T_e$. The top panel shows the initial (thermal equilibrium) state, where all subbands have a Maxwellian distribution with the extractor having the highest occupation. At time $t = 2.5$ ps, the in-plane distribution has heated considerably for all subbands, with the lower lasing level being hottest at $T_e = 136$ K. At $t = 10$ ps, the lower lasing level has cooled down while the other states have heated up, and the injector state has the highest temperature of 220 K. At $t = 100$ ps, the system has reached a steady state, where the lower lasing level is coolest (92 K) and the injector is hottest (203 K). A noticeable feature in Fig. 8 is the big difference in temperature of the different subbands, with a temperature difference of 110 K between the injector and lower lasing level. In addition to having very different temperatures, the in-plane energy distributions are very different from a heated Maxwellian. A weighted average (using occupations as weights) of the steady-state electron temperatures is 159 K, which is 109 K higher than the lattice temperature.

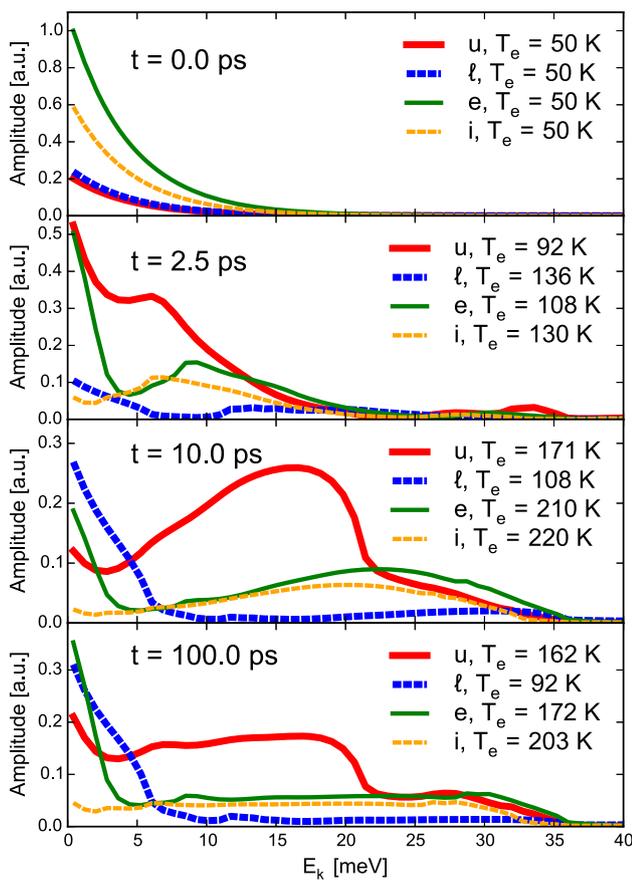


Fig. 8 In-plane energy distribution for all eigenstates shown in Fig. 1, except for the high energy state (*h*). Results are shown for four values of time, starting in thermal equilibrium ($t = 0$ ps) and ending in the steady state ($t = 100$ ps). Also shown are the corresponding electron temperatures, calculated from $T_e = \langle E_k \rangle / k_B$

5 Conclusion

We derived a Markovian master equation (33) for the single-electron density matrix, including off-diagonal matrix elements (coherences) as well as in-plane dynamics. The MME conserves the positivity of the density matrix, and accounts for scattering of electrons with phonons and impurities. We applied the MME to simulate electron transport in a THz QCL. Close to lasing (around the design electric field), our results for current density are in good agreement with both experiment and theoretical results based on NEGF. The differences between NEGF and density matrix at low fields are small and can be attributed to the omission of interface roughness scattering in our simulation and the effects of collisional broadening.

We have shown that the magnitude of the off-diagonal density-matrix elements can be a significant fraction of the largest diagonal element. With the device biased for lasing, the greatest coherence was between the injector and extractor levels, with a magnitude of 21 % of the largest diagonal

element (the upper lasing level). This result demonstrates the need to include coherences when describing QCLs in the THz range.

We have found that significant electron heating takes place at the design electric field, with in-plane distributions deviating far from a heated Maxwellian distribution. The electron temperature was found to vary strongly between subbands, with an average subband temperature about 109 K hotter than the lattice temperature of 50 K. This result demonstrates the need to treat in-plane dynamics in detail.

Time-resolved results showed that, early in the transient, the current density exhibits high-amplitude coherent oscillations with a period of 100–200 fs, decaying to a constant value on a time scale of 3–10 picoseconds. The amplitude of current oscillations could be over 10 times larger than the steady-state current. Occupations of subbands and in-plane energy distributions took considerably longer (20–100 ps) than the current density to reach the steady state.

Solving the MME for the density matrix is a numerically efficient approach to time-dependent quantum transport in nanostructures far from equilibrium.

Acknowledgments The authors gratefully acknowledge the support provided by the U.S. Department of Energy, Basic Energy Sciences, Division of Materials Sciences and Engineering, Physical Behavior of Materials Program, Award No. DE-SC0008712. The work was performed using the resources of the UW-Madison Center for High Throughput Computing (CHTC).

Appendix: Calculation of \mathcal{G} terms

This appendix is devoted to explicit calculation of \mathcal{G} . This task involves the evaluation of Eq. (21) for different scattering mechanisms, which is repeated here for convenience:

$$\mathcal{G}_g^{\text{em}}(E_k, Q_z, E_q) \equiv \frac{1}{2\pi} \int_0^{2\pi} d\theta \mathcal{W}_g^{\text{em}}(\mathbf{Q}, E_k). \quad (44)$$

The matrix element $\mathcal{W}_g^{\text{em}}(\mathbf{Q}, E_k)$ can always be written in terms of $E_q = \hbar^2 q^2 / 2m^*$, $E_k = \hbar^2 k^2 / 2m^*$, $E_z = \hbar^2 Q_z^2 / 2m^*$, and the angle θ between \mathbf{k} and \mathbf{q} . Note that this definition of E_z is only to make expressions more compact and readable, and the actual energy in the z -direction is contained in the Δ_{nm} terms. Derivations of the various phonon matrix elements $\mathcal{W}_g^{\text{em}}(\mathbf{Q})$ used in this section can be found in Refs. [26, 27].

For the case of longitudinal acoustic (LA) phonons, we employ the equipartition approximation and get

$$\mathcal{W}_{\text{LA}}^{\text{em}}(\mathbf{Q}) \simeq \frac{D_{\text{ac}}^2 k_B T_L}{2m^* v_s^2} = \beta_{\text{LA}}, \quad (45)$$

where D_{ac} is the deformation potential for acoustic phonons, and v_s is the sound velocity in the material. In this case, the θ integration in Eq. (44) gives $\mathcal{G}_{LA}^{em} = \beta_{LA}$. Since acoustic phonons are treated elastically, the emission and absorption terms are identical.

As with the acoustic phonons, the nonpolar optical phonon scattering is isotropic, so the phonon matrix element is constant $\mathcal{W}_{op}^{em}(\mathbf{Q}) = (N_{op} + 1)\beta_{op}$. The angular integration in Eq. (44) gives $\mathcal{G}_{op}^{em} = (N_{op} + 1)\beta_{op}$.

The phonon matrix element for electron scattering with polar optical phonons, with screening included, is given by

$$\mathcal{W}_{pop}^{em}(\mathbf{Q}) = (N_{pop} + 1)\beta_{pop} \frac{Q^2}{(Q^2 + Q_D^2)^2}, \tag{46a}$$

where Q_D is the Debye wave vector defined by $Q_D^2 = ne^2/(\epsilon k_B T)$ and

$$\beta_{pop} = \frac{e^2 E_{pop}}{2\epsilon_0} \left(\frac{1}{\epsilon_r^\infty} - \frac{1}{\epsilon_r} \right), \tag{46b}$$

where ϵ_r^∞ and ϵ_r are the high-frequency and low-frequency relative permittivities of the material, respectively, and n is the average electron density. The effects of screening are quite small at the electron densities considered in this work; however, the $1/Q^2$ singularity poses problems in numerical calculations due to the high strength of the POP interaction. These problems are avoided by including screening. We can now calculate

$$\begin{aligned} & \frac{\mathcal{G}_{pop}^{em}(E_k, E_z, E_q)}{(N_{pop} + 1)} \\ &= \frac{\beta_{pop}}{2\pi} \int_0^{2\pi} d\theta \frac{|\mathbf{Q} - \mathbf{k}|^2}{(|\mathbf{Q} - \mathbf{k}|^2 + Q_D^2)^2} \\ &= \frac{\beta_{pop}}{2\pi} \int_0^{2\pi} d\theta \frac{Q_z^2 + q^2 + k^2 - 2qk \cos(\theta)}{(q_z^2 + q^2 + k^2 - 2kq \cos(\theta) + Q_D^2)^2} \\ &= \frac{\beta_{pop}}{Q_z^2 + q^2 + k^2} \left[1 + \frac{Q_D^2}{Q_z^2 + q^2 + k^2} - \frac{4k^2 q^2}{(Q_z^2 + q^2 + k^2)^2} \right] \\ &\quad \times \left[\left(1 + \frac{Q_D^2}{Q_z^2 + q^2 + k^2} \right)^2 - \frac{4k^2 q^2}{(Q_z^2 + q^2 + k^2)^2} \right]^{-\frac{3}{2}} \\ &= \frac{\hbar^2 \beta_{pop}}{2m^* E_z + E_k + E_q} \\ &\quad \times \left[1 + \frac{E_D}{E_z + E_q + E_k} - \frac{4E_k E_q}{(E_z + E_q + E_k)^2} \right] \\ &\quad \times \left[\left(1 + \frac{E_D}{E_z + E_q + E_k} \right)^2 - \frac{4E_k E_q}{(E_z + E_q + E_k)^2} \right]^{-\frac{3}{2}}, \tag{47} \end{aligned}$$

where $E_D = \hbar^2 Q_D^2 / 2m^*$ is the Debye energy.

The matrix element for ionized impurities is given by

$$\mathcal{W}_{ii}^{em}(\mathbf{Q}) = \frac{\beta_{ii}}{|\mathbf{Q}|^4} \tag{48a}$$

with

$$\beta_{ii} = \frac{N_I Z^2 e^4}{2\epsilon_r^2 \epsilon_0^2}, \tag{48b}$$

where N_I is the impurity density, and Z is the number of unit charges per impurity. This matrix element gives

$$\begin{aligned} & \mathcal{G}_{ii}^{em}(E_k, E_z, E_q) \\ &= \frac{\beta_{ii}}{2\pi} \int_0^{2\pi} d\theta \frac{1}{(Q_z^2 + q^2 + k^2 - 2qk \cos(\theta))^2} \\ &= \beta_{ii} \frac{Q_z^2 + q^2 + k^2}{[(Q_z^2 + q^2 + k^2)^2 - 4q^2 k^2]^{3/2}} \\ &= \beta_{ii} \frac{\hbar^4}{4m^2} \frac{E_z + E_k + E_q}{[(E_z + E_k + E_q)^2 - 4E_k E_q]^{3/2}}. \tag{49} \end{aligned}$$

Since ionized-impurity scattering is elastic, the absorption term is identical to the emission term.

References

1. Faist, J., et al.: Quantum cascade laser. *Science* **264**, 553 (1994)
2. Dupont, E., Fatholouloumi, S., Liu, H.: Simplified density-matrix model applied to three-well terahertz quantum cascade lasers. *Phys. Rev. B* **81**, 205311 (2010)
3. Jirauschek, C., Kubis, T.: Modeling techniques for quantum cascade lasers. *Appl. Phys. Rev.* **1**, 1, 011307 (2014)
4. Iotti, R., Rossi, F.: Nature of charge transport in quantum-cascade lasers. *Phys. Rev. Lett.* **87**, 146603 (2001)
5. Callebaut, H., et al.: Importance of electron-impurity scattering for electron transport in terahertz quantum-cascade lasers. *Appl. Phys. Lett.* **84**, 5, 645 (2004)
6. Gao, X., Botez, D., Knezevic, I.: X-valley leakage in gaas-based midinfrared quantum cascade lasers: a monte carlo study. *J. Appl. Phys.* **101**, 6, 063101 (2007)
7. Shi, Y.B., Knezevic, I.: Nonequilibrium phonon effects in mid-infrared quantum cascade lasers. *J. Appl. Phys.* **116**, 12, 123105 (2014)
8. Willenberg, H., Döhler, G.H., Faist, J.: Intersubband gain in a bloch oscillator and quantum cascade laser. *Phys. Rev. B* **67**, 085315 (2003)
9. Kumar, S., Hu, Q.: Coherence of resonant-tunneling transport in terahertz quantum-cascade lasers. *Phys. Rev. B* **80**, 245316 (2009)
10. 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**, 165322 (2009)
11. Terazzi, R., Faist, J.: A density matrix model of transport and radiation in quantum cascade lasers. *N. J. Phys.* **12**, 3, 033045 (2010)
12. Lee, S.-C., Wacker, A.: Nonequilibrium green's function theory for transport and gain properties of quantum cascade structures. *Phys. Rev. B* **66**, 245314 (2002)

13. Callebaut, H., Hu, Q.: Importance of coherence for electron transport in terahertz quantum cascade lasers. *J. Appl. Phys.* **98**, 10, 104505 (2005)
14. Dupont, E., et al.: A phonon scattering assisted injection and extraction based terahertz quantum cascade laser. *J. Appl. Phys.* **111**, 7, 073111 (2012)
15. Lindskog, M., et al.: Comparative analysis of quantum cascade laser modeling based on density matrices and non-equilibrium green's functions. *Appl. Phys. Lett.* **105**, 10, 103106 (2014)
16. Harrison, P., Indjin, D.: Electron temperature and mechanisms of hot carrier generation in quantum cascade lasers. *J. Appl. Phys.* **92**, 11, 6921 (2002)
17. Frohlich, H.: Theory of electrical breakdown in ionic crystals. *Proc. R. Soc. Lond. A* **160**, 220 (1937)
18. Breuer, H.P., Petruccione, F.: *The Theory of Open Quantum Systems*. Oxford University Press, Oxford (2002)
19. Knezevic, I., Novakovic, B.: Time-dependent transport in open systems based on quantum master equations. *J. Comput. Electron.* **12**, 3, 363 (2013)
20. Kohn, W., Luttinger, J.M.: Quantum theory of electrical transport phenomena. *Phys. Rev.* **108**, 590 (1957)
21. Fischetti, M.V.: Master-equation approach to the study of electronic transport in small semiconductor devices. *Phys. Rev. B* **59**, 4901 (1999)
22. Karimi, F., Davoody, 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)
23. Mutze, U.: An asynchronous leapfrog method ii (2013) (**Unpublished**). [arXiv:1311.6602](https://arxiv.org/abs/1311.6602)
24. Wacker, A.: Semiconductor superlattices: a model system for non-linear transport. *Phys. Rep.* **357**, 1, 1 (2002). (**ISSN 0370-1573**)
25. Moiseyev, N.: Quantum theory of resonances: calculating energies, widths and cross-sections by complex scaling. *Phys. Rep.* **302**, 5-6, 212 (1998). (**ISSN 0370-1573**)
26. Ferry, D.K.: *Semiconductors*. IOP Publishing, Bristol (2013). (**ISBN 978-0-750-31044-4**)
27. Jacoboni, C., Lugli, P.: *The Monte Carlo Method for Semiconductor Device Simulation*. Springer, Vienna (1989)
28. Iotti, R., Ciancio, E., Rossi, F.: Quantum transport theory for semiconductor nanostructures: a density-matrix formulation. *Phys. Rev. B* **72**, 125347 (2005)