RESEARCH ARTICLE | MARCH 18 2024

# Combined resonant tunneling and rate equation modeling of terahertz quantum cascade lasers

Zhichao Chen ; Andong Liu ; Dong Chang ; Sukhdeep Dhillon ; Manijeh Razeghi ; Feihu Wang



*J. Appl. Phys.* 135, 115703 (2024)
https://doi.org/10.1063/5.0198059

Journal of Applied Physics

ARTICLE

pubs.aip.org/aip/ja p

# Combined resonant tunneling and rate equation modeling of terahertz quantum cascade lasers

Cite as: J. Appl. Phys. 135, 115703 (2024); doi: 10.1063/5.0198059
Submitted: 16 January 2024 · Accepted: 27 February 2024 ·
Published Online: 18 March 2024



Zhichao Chen,[1,a)] Andong Liu,[1] Dong Chang,[2] Sukhdeep Dhillon,[3] and Feihu Wang[1,a)] Manijeh Razeghi,[4]

AFFILIATIONS [1] Shenzhen Institute for Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China [2] International Quantum Academy, Shenzhen 518048, China
[3] Laboratoire de Physique de l'Ecole Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, 75014 Paris, France [4] Center for Quantum Devices, Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, Illinois 60208, USA

[a)] Authors to whom correspondence should be addressed: chenzhichao1994@hotmail.com and wangfh@sustech.edu.cn



## ABSTRACT

Terahertz (THz) quantum cascade lasers (QCLs) are technologically important laser sources for the THz range but are complex to model. An efficient extended rate equation model is developed here by incorporating the resonant tunneling mechanism from the density matrix formalism, which permits to simulate THz QCLs with thick carrier injection barriers within the semi-classical formalism. A self-consistent solution is obtained by iteratively solving the Schrödinger–Poisson equation with this transport model. Carrier–light coupling is also included to simulate the current behavior arising from stimulated emission. As a quasi-ab initio model, intermediate parameters, such as pure dephasing time and optical linewidth, are dynamically calculated in the convergence process, and the only fitting parameters are the interface roughness correlation length and height. Good agreement has been achieved by comparing the simulation results of various designs with experiments, and other models such as density matrix Monte Carlo and non-equilibrium Green's function method that, unlike here, require important computational resources. The accuracy, compatibility, and computational efficiency of our model enable many application scenarios, such as design optimization and quantitative insights into THz QCLs. Finally, the source code of the model is also provided in the supplementary material of this article for readers to repeat the results presented here, investigate, and optimize new designs.



## I. INTRODUCTION

Terahertz (THz) quantum cascade lasers (QCLs), demonstrated two decades ago,[1] are currently the most advanced semiconductor lasers in the spectral range between 1 and 5 THz. With proven potential in many applications, including security screening, remote sensing, nondestructive imaging, high-speed communications, astronomy, biology, and medicine, THz QCLs are believed to be the underpinning devices for the upcoming THz technology revolution.[2–4] So far, the highest record for pulsed THz QCL operation is 261 K (−12 °C).[5] However, their operation at room temperature is still an unresolved challenge. The temperature performance of QCLs is related to many complex interconnected physical mechanisms, resulting in the difficulty in QCL design and optimization, which has led to a very slow improvement of operation temperature in the past decades.[6] To accelerate the development of THz QCLs, an efficient and accurate design/optimization tool is needed to push the limit of THz QCLs.

Despite the success in the accuracy of the density matrix (DM) method[7–17] and the non-equilibrium Green's function (NEGF) approach,[18–24] computational speed is always crucial for design and optimization tasks of complex structures like QCLs. Progress has been made in the past decade on DM, for example, simplified DM incorporated with the Monte Carlo algorithm[15] and improved infinite-period DM approach without the need for the a priori defined dominant transport states.[11] Also, a completely positive Markovian evolution of DM is demonstrated;[12–14] the time dependency and full in-plane dynamics captured in such a method provide a deep insight into the physics process in QCLs. Advanced DM, derived from first

principle,[12–14,16,17] enables researchers to study more general cases. Differing from DM in, e.g., Refs. 7, 9, and 15, well-defined eigenstates are used instead of manually chosen tight-binding states separated by one or more coupling barriers. Nevertheless, such a model usually has higher numerical and mathematical complexity to be implemented. The quantum mechanical NEGF approach has been proven to be a powerful tool to investigate, design, and optimize QCL devices,[18,25–27] including those based on commercialized software, e.g., nextnano.[5,24,28–30] However, as NEGF can easily involve a lot of computational resources, it is difficult to run on computers other than the modern high-performance computing (HPC) clusters within a reasonable time. Compared with the above-mentioned two approaches, scattering-based hopping transport models like Monte Carlo and self-consistent rate equation methods are less computationally demanding. For Monte Carlo, a large ensemble of carriers is considered (typically $10^4$–$10^5$) in a stochastic way. With each carrier trajectory tracked and updated in many short time intervals, the system evolves in the time domain and finally reaches its steady state. The 3D Monte Carlo method explicitly considers the intrasubband scattering processes and makes the in-plane electron distribution form automatically without making any thermalization assumption. Though the Monte Carlo simulation provides a deeper insight into the physical nature of QCL, the complexity is much higher than that of the rate equation model. By applying similar conditions, simulation results show that the in-plane carrier distribution by Monte Carlo is indeed very close to the Fermi–Dirac function,[31,32] which can be directly adopted to describe the in-plane carrier distribution and simplify the 3D problem to be 1D. The solution of the rate equations still needs a self-consistent algorithm because of the a priori unknown carrier distribution and scattering rates.[33] The convergence speed and numerical stability are significantly improved by dynamically changing the weighting factor.[34] Despite efforts devoted to simplifying the DM and NEGF, for example, by neglecting the in-plane wavevector dependency to reduce the order of the density matrix[7,9] and allowing the scattering self-energies to be k-independent,[35,36] the self-consistent rate equation method is still the most computationally efficient compared to the other methods. Here, we will show that the resonant tunneling transport mechanism incorporated self-consistent rate equation model developed in this work has comparable accuracy with other QCL modeling techniques. Beyond the above-mentioned selfconsistent model, the reduced rate equation (RRE) model[37–40] considering only a subset [e.g., the upper and lower lasing level (LLL) lifetime] of laser parameters is a further simplification of the full self-consistent RE model. Taking advantage of its simplicity, RRE models are usually used to study the dynamic behavior of QCLs. The previous RRE model treats the laser parameters as constant, making such a model only valid around designed bias or certain temperatures.[37,38] Recently, the RRE model has been extended by incorporating the bias and temperature-dependent parameters extracted from the full self-consistent RE model.[39,40] As the semiclassical RE model is still used to obtain the input parameters of the RRE model in Ref. 39, our model, where the drawback of the semi-classical method has been overcome, could help to further improve the accuracy of the current REE models.

Semiclassical models have been frequently used in QCL design and modeling at the early stage. Both Monte Carlo and rate equation calculations have been compared with experiments, showing adequate validity.[33,34,41–45] However, the semiclassical method only considers the incoherent scattering mechanism, with which the whole structure of the simulation window (typically three QCL periods) is considered by a single and well-defined Hamiltonian. The fundamental limitation of such a method, discussed by many references,[15,46,47] can be easily found, especially in THz QCL modeling, where a thicker injection barrier (~5 nm) is usually used to suppress the wrong injection channel for a small photon transition energy (~10 meV). Consequently, the anti-cross energy gap corresponding to the coupling energy is very small, and, thus, the electron transport is dominated by resonant tunneling in this case. In the semiclassical method, however, the quantum coherent tunneling and dephasing are neglected, and the transition rates depend only on the scattering-induced process. At alignment bias, the wavefunctions of the two states extend across the barrier, and the instantaneous event opens a "short-cut" for the electrons transport across the barrier, and an unphysical spike of the current density will appear in the calculation. Recently, coherent evolution, which is naturally included in DM, has been incorporated into the existing semiclassical framework to describe the carrier transport across a thick barrier in THz QCLs. Callebaut and Hu[46] first included coherent transport in the Monte Carlo model. Instead of using full DM, the intra-module was treated by the semiclassical MC approach, and the transport through the barrier was handled by solving the Liouville equation. The pure dephasing time is, however, from a phenomenological constant value for all subbands. Later, Jirauschek[15] further improved and simplified the framework of DM-MC. In his model, instead of treating the Boltzmann transport equation (MC) and Liouville equation (DM) simultaneously, the intra-module transition rate was described by the tunneling rate equation. Hence, this model is closer to the "hopping transport" model and more compatible with the MC framework. Another improvement from Jirauschek is that the dephasing time is not from the phenomenological value but calculated according to the intra-subband scattering rate by Ando's model,[48,49] providing a more ab initio way of simulation. Apart from the MC, the rate equation method has also been extended by including the tunneling rate from the DM formalism. Two typical models are from Terrzi[47] and Razavipour.[50] Researches based on these enhanced rate equation methods have shown great potential in modeling THz QCLs.[51,52] However, further improvement is needed for these models. For example, empirical values of the pure dephasing time and optical linewidth are still used.[50,51] However, in Ref. 52, the pure dephasing rate is calculated based on Ando's model, the carrier–light coupling is neglected, resulting in inaccurate current density estimation after lasing. Terazzi's model is a rather ab initio model without additional fitting parameters, but to our knowledge has only been applied to mid-infrared (MIR) QCLs,[47] where the operational physics are significantly different. In our model, some modifications have been applied and accuracy has been improved. For example, the coupling strength calculated by the method in Terrazi's model has been found

underestimated by around 20%. We use an alternative way for the coupling strength and the result is improved when compared with the exact value of the anticrossing gap from a well-defined Hamiltonian. Moreover, electron– electron (EE) scattering, which could be essential to modeling THz QCL, is also neglected in Terazzi's model.

Based on the framework of Terazzi's model[47] and taking advantage of MC and rate equation methods,[15,50–54] we aim to demonstrate a comprehensive self-consistent rate equation model of THz QCLs. One or more missing effects (as mentioned in the last paragraph) in the existing rate equation model with similar configuration have been complemented.[47,50–52] The missing effects in these literatures also bring about some problems to catch the experimental result. For example, in Ref. 51, as dephasing time varies with bias, different fitting values have to be applied to match the experiment I–V curve. Although in Ref. 52, improvements have been made by including the leakage to continuum and calculating the dephasing rate with Ando′s model, dispensing with the empirical input. However, without carrier–light coupling, the discrepancy cannot be compensated by their leakage model after lasing threshold. EE scattering can be important and have a significant impact on certain structures.[34,55] Hence, it is important to include all these effects in a single model. Some major parts of the simulation or techniques are carefully selected and also different from the literature above. Detailed investigation and comparison for these changes, for example, tunneling coupling strength and secondorder current have been given. Our model follows the spirit of ab initio modeling. The only structural fitting parameters in the transport model are the effective interface roughness height and correlation length. All other intermediate parameters, such as the pure dephasing time and optical linewidth, are dynamically calculated.

This paper is organized as follows: Sec. II discusses the theoretical basics and technical details of the model. Section III will demonstrate the calculation results using our model and compare them with experimental results. Last, a summary of the study and some potential improvements of the model will be discussed in Sec. IV.

## II. THEORY AND MODEL

Some key features and modifications have been applied based on Terazzi's work to tailor our model to satisfy the requirements of THz QCL design.[9,47] (1) An improved equation [Eq. (13)] is used to calculate the coupling strength rather than the "first-order approximation" proposed in Terazzi's model.[47] The latter uses the localized wavefunction and effective mass profile to compute the coupling strength. However, we found that the method used in this work (using localized potential instead of effective mass) has a better match with the anticrossing gap obtained by the well-defined potential profile of symmetric two-well structures. Some additional comparison of these two methods is presented in Sec. II C. (2) EE scattering is included. It is argued that the EE process involves four states, and the $N^4$ complexity (N is the number of states) makes it computationally expensive.[32] However, in most THz QCLs, only a few states are involved in the transport process. Thus, it is reasonable to consider EE scattering for these states, and due to the commutability, the most time-consuming calculation process of the form factor can be simplified by skipping those combinations with identical results. (3) Instead of calculating the dephasing rate at thermal energy ($k_{th}$ ¼ $\hbar^{1}p2m_{k}k_{B}T_{e}ffi$, where $k_B$ is the Boltzmann constant and $T_e$ is the electron temperature), we compute the dephasing rate in the whole k-space and averaged them according to the Fermi–Dirac distribution before entering the 1D transport model. (4) Both of the first- and second-order approximations of the current density are considered. However, it seems the secondorder current developed in Ref. 47 underestimated the current before alignment at low temperatures. Hence, we retain the firstorder current approximation since it has proven to work well for QCLs.[15] (5) Instead of using the method in Ref. 47, where the photon population is numerically converged, the carrier–light coupling is carried out based on the time-evolution of classical light intensities from the MC model. The latter is more straightforward and compatible with our rate equation model without significantly increasing the computational load. A similar time-evolution behavior of the light intensity as in Ref. 54 is made. Other features such as nonparabolicity, self-self-consistent Schrödinger–Poisson equation, and kinetic energy balancing (electron temperature calculation) are also included in our model.

### A. Bandstructure

The tunneling rate of carriers in QCL is derived from the DM formalism, where the band structure needs to be calculated with the tight-binding Hamiltonian.[32,53] In this model, the periodic active region structure is separated by the injection barriers, as the latter usually acts as the "bottlenecks" of current circulation in THz QCLs. Each period is a single module containing the localized basis states. The tight-binding potential $V_{tb}$ is then defined for each identical module. Note that because we use the matrix solution of the Schrödinger equation[32] to ensure the wavefunction decays properly to zero at the quasi-infinite edge of the simulation window, the most left and right barriers must be numerically extended thick enough according to the injection barrier potential and the applied electric field since we need two periods to close the system for the establishment of rate equations. The solutions of the adjacent module are then duplicated and shifted by the period length and bias per period. Next, nonparabolicity is included by considering the energy dependent quantization effective mass in the Schrödinger equation

$$\frac{\hbar^2}{2}\frac{d}{dz}\left(\frac{1}{m^*(E,z)}\frac{d\psi(z)}{dz}\right)+V(z)\psi(z)=E\psi(z), \quad (1)$$

where V(z) is the potential profile, which is equal to the tightbinding potential $V_{tb}$ here, E is the eigenenergy, and ψ(z) is the envelope wavefunction of electrons. The energy dependent quantization effective mass $m^*(E, z) = m^*(z)[1 + \alpha^0(z)(E - V(z))]$, and the nonparabolicity parameter $\alpha^0 = (E_g + \Delta_{so}/3)$, where $E_g$ is the bandgap

and $\Delta_{so}$ is the split-off energy. The non-linear eigenvalue problem of Eq. (1) is then solved by a matrix finite difference method (FDM).[56,57] For the in-plane effective mass of each subband, nonparabolicity can be easily considered by averaging z and E dependent effective mass according to the wavefunction. However, for clarity and simplicity, we assume the in-plane isotropy and take the in-plane effective mass as a constant of GaAs (0:067$m_0$). Though the in-plane nonparabolicity can cause effects like additional optical broadening, it is found to be negligible in GaAs-based THz QCLs because of a large bandgap of GaAs material system, and the main states involved in the carrier transport process lie on the bottom of the quantum wells.[58] Last, to deal with the nonparabolicity, the k.p method considering the coupling between the conduction band and the valence band is also used in this field.[14,59]

The Hartree approximation of electron–electron interactions is implemented by solving Eq. (1) together with the Poisson equation

$$\frac{d}{dz}\left(\epsilon\frac{dV}{dz}\right) = q_e^2\left(n_D(z) - \sum_i n_i^s|\psi_i(z)|^2\right), \quad (2)$$

where $\epsilon(z) = \epsilon_0\epsilon_r$ is the permittivity, $V_c(z)$ is the electrostatic (self) potential, $q_e$ is the elementary charge, $n_D(z)$ is the volume doping density, and $n_i^s$ is the sheet carrier density of state i. The Schrödinger–Poisson system can be solved based on the thermal distribution (Fermi–Dirac) with the minimum computational efforts.[60,61] A more accurate way is to determine the carrier distribution by the carrier transport model,[53] but it is time-consuming since the transport model needs to be invoked several times until the Schrödinger–Poisson system converges. Benefiting from the fast calculation speed of the rate equation method, the computational time in this model is in a reasonable range (typically within 1 min with ~5 S–P iterations). It is essential to note the periodicity of the charge density, where Eq. (2) is strictly resolved within a single period length $[z_0, z_0 + L_p)$, and $z_0$ is the starting coordinate of the period. The localized wavefunction itself, however, does not fulfill the periodic condition as the up- and downstream module wavefunction could extend to the central one. Thus, the periodicity of the probability density is established by adding the tails of the wavefunction in the adjacent modules into the central one, then $|\psi_i|^2 = |\psi_{i1}|^2 + |\psi_{i+1}|^2$ and $|\psi_{in}(z)|^2 = |\psi_i(z + nL_p)|^2$.

The boundary condition is set by $V_c(z_0) = V_c(z_0 + L_p) = 0$, and it is again solved by the finite difference method. Finally, the total potential $V = V_0 + V_c$, where $V_0$ is the potential without considering the space charge effects, is again entering Eq. (1), and the whole process run iteratively until convergence is achieved. The electrostatic potential of the other region in the simulation window, say the right period and the extended barriers, can be obtained by shifting and duplicating from the central period. Figure 1 shows the self-self-consistent result of the S–P system. As for the THz QCL,[62] the carrier sheet density is $3 \times 10^{10}$ cm$^2$, the magnitude of the selfpotential is less than 1 meV, as can be seen in Fig. 1(a). And the influence on the potential is barely visible. Another case for midinfrared (MIR) QCL[63] is given in Fig. 1(b) to illustrate the intuitive relation between the effect of the S–P equation and the doping density. Here, the doping density is much higher, causing the sheet carrier density per period to be about one order of magnitude larger. The magnitude of the self-potential is also roughly ten times larger than that in Fig. 1(a), resulting in significant bending of the potential profile. For this reason, the S–P equation is usually included in the modeling of MIR QCLs,[9,25,64] but sometimes ignored in THz QCLs.[50–52] Thus, with low doping density, the S–P equation may be disregarded to maintain high computational efficiency in design and optimization processes. Furthermore, other phenomenological effects such as interdiffusion[32,65] could bring much larger change to the potential profile than the S–P effect. To maintain the clarity, this effect is not included in our model.

### B. Scattering mechanisms

In this study, five scattering mechanisms are considered for the intra-module transport. They are longitudinal optical phonon (LO) scattering, EE scattering, impurity (IMP) scattering, interface roughness (IFR) scattering, and alloy disorder (AD) scattering. Because the localized states within a module are strongly coupled, and dephasing is less important, resonant tunneling is disregarded inside the module, and only inter-subband scattering needs to be considered. From Fermi's golden rule, the scattering rate of a specific mechanism (m), from state $j_i$, $k_i$ to all possible state in the final subband j reads

$$W_{ij(m,k)_i} = \frac{2\pi}{\hbar}\sum_{k_j} |\langle \psi_{j,k_j}|H'_{(m)}|\psi_{i,k_i}\rangle|^2 \delta(E_{j,k_j} - E_{i,k_i}), \quad (3)$$

with $H'$ is the perturbing Hamiltonian causing scattering, k is the wavevector, and $E_{i,k_i}$ is the energy of the state $j_i$, $k_i$. Note the above equation is for the elastic process. For the inelastic process (e.g., phonon scattering), the delta function becomes $\delta(E_{j,k_j} - E_{i,k_i} \pm \hbar\omega_0)$, where the upper sign (+) stands for emission, the lower sign (−) stands for absorption, and $\hbar\omega_0$ is the phonon energy. We do not show the explicit equations for every scattering mechanism for clarity of the paper. Detailed information about the scattering equations, the symbols, and the references are given in Table I in Appendix A. The mean scattering rate is computed according to the Fermi–Dirac distribution before entering the rate equations[32]

$$W^{(m)}_{ij,k_i}(k_i) = \int [1 - f_j(k_j)]\, dk_j$$

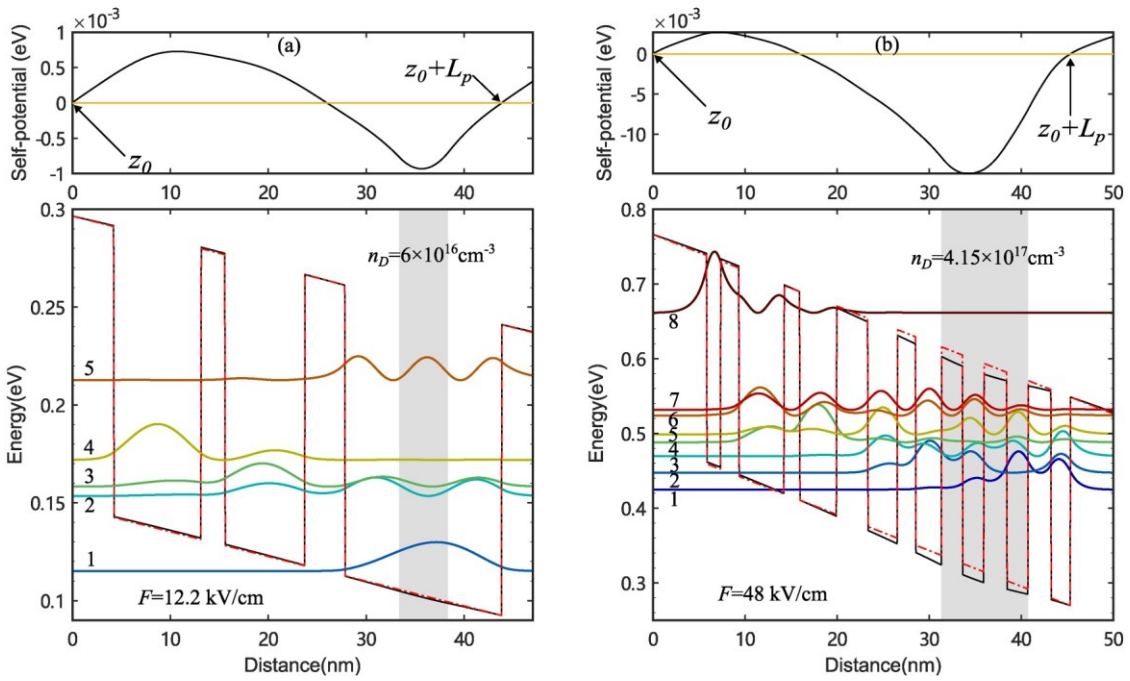



FIG. 1. Solution of the Schrödinger–Poisson equation. The self-consistent potentials are displayed in the top row. The original potential of the material band edge (dashed red) and the bent potential (solid black), together with the converged wavefunctions, are shown in the bottom row. The shaded areas are the doping region. (a) A THz QCL with electron sheet density per period of 3 $10^{10}$ cm$^2$.[62] (b) A MIR QCL with electron sheet density per period of 3:9 $10^{11}$ cm$^2$.[63]

$$W_{ij} = \frac{}{} \pi n_{si} \quad : \quad (4)$$

In the above equation, k ¼ jkj, and the carrier distribution $f_i(k_i)$ is the Fermi–Dirac function in this model. Note that $f_i(k_i)$ is also a function of the quasi-Fermi level in each subband $\mu_i$, It is related to the 2D carrier density $n^s_i$ of each state by the equation

$$n_i = \frac{m_{ki}}{\pi \hbar^2} k_B T_e \ln\left(1 + e^{\frac{\mu_i - E_{i,0}}{k_B T_e}}\right), \quad (5)$$

where $\mu_i$ is then solved by bisection method dynamically according to the $n^s_i$ in the current iteration. The final state blocking arising from Pauli's exclusion principle is incorporated by including $[1 - f_j(k_j)]$ in the integral. The numerical upper limit of the wavevector for all subbands is defined as $k_{max} = \sqrt{2m_k(V_{max} - E_{1,0})/\hbar^2}$, $V_{max}$ is the maximum potential within a module and $E_{1,0}$ is the first subband energy at k = 0. In particular, EE scattering is a two-body problem that involves four states. The scattering rate between two states is calculated by considering all possible combinations of the subbands involved, i.e., $W_{ij(EE)} = P_0 W_{ii(EE_{0jj_0})} + W_{iijj(EE)} - W_{ijji(EE)}$.[32] Final state blocking is dismissed when averaging the EE scattering rate for simplification. The total scattering rate is calculated by summing up all scattering mechanisms $W_{ij} = \sum_m P_m W_{ij}^{(m)}$.

It is important to mention that the cut-off $E_{cut}(k)$ energy exists in the scattering rate.[32] For LO phonon scattering, this is because the energy difference between the initial and final subbands is too large or too small for states with a small k-vector to absorb or emit an LO phonon, where the scattering cannot occur. For the elastic process, the scattering from a lower subband to a higher subband cannot happen if the kinetic energy of the lower subband does not exceed the bottom of the higher subband as the energy conservation has to be satisfied. Then, the scattering rate below this cut-off wavevectorthe integration in Eq. (4) should start from the cut-off wavevector.$k_{cut} = \sqrt{2m_k E_{cut}/\hbar^2}$ is all zero, and

We uniformly apply the Debye and Thomas–Fermi approximation in the screening of the LO phonon scattering, EE scattering, and IMP scattering,[66] where the inversion screen length is calculated by

$$q_{2s} = \begin{cases} \dfrac{3e^2 n_{3D}}{n^{3}_{avg}{}^{D} E_{avgFe2}}, & (k_B T < 2E_F/3), \\ \epsilon k_B T_L, & (k_B T > 2E_F/3), \end{cases} \tag{6}$$

with the mean 3D electron density $n^{3}_{avg}{}^{D}$ and Fermi energy for electron gas at 0 K $E_F = (3\pi^2 n_e h^3)^{2/3}/2m^*$. In the above equation, Debye's model is used for high temperatures, as the latter fails at low temperatures, the Thomas–Fermi model is then used to compensate such a scenario. This approach has been proven to be accurate and computationally efficient compared to the full random phase approximation (RPA)[67] as the 3D electron density and lattice temperature in Eq. (5) do not change on the run of self-consistent iteration and the kinetic energy balance iteration. Thus, the precalculated k-dependent scattering rate can be reused without recalculating them in every self-consistent iteration. Though advanced methods like full RPA[68] and simplified 2-D RPA[66] provide better results, they will relate the screening effect to the subband carrier distribution, leading to the recalculation of the scattering rate at each self-consistent iteration and, therefore, considerably slowing down the convergence speed.

The pure dephasing contribution to the broadening of the tunneling rate and optical transitions are accounted for by the intra-subband scattering from Ando's model.[48,49] As EE scattering is explicitly excluded in Ando's model when identical effective mass is used for each subband,[48] and the contribution of LO phonon scattering is negligible,[15,49,69] only the elastic processes (IMP, IFR, AD) are considered for pure dephasing rates. The pure dephasing rate is given by

$$\gamma^{*}_{ij,m_k} = \pi X_k D \sim^{(m)} ikE \; D\, jkojH\sim^{(m)} jjkE^2 \delta(\epsilon_k - \epsilon_{k^0}), \gamma$$

$$= h_0 ikojH\, j \tag{7}$$

where $\epsilon_k = h^2k^2/(2m^k)$. Again, the in-plane nonparabolicity is ignored in this model, and the above equation only holds for identical effective mass in two subbands. More information can be found in Ref. 47 when nonparabolicity and nonuniform in-plane effective mass are used. The phenomenological value of the pure dephasing time for all subbands, with the relation $\tau^{*}_{ij} = (\gamma^{*}_{ij})^{1}$, are sometimes used to fit the experiment data, typically around 0.3–1 ps.[7,50,51] In this work, however, the pure dephasing rate is not a fitting parameter but is calculated using Eq. (6). The source of the explicit equations for different scattering mechanisms again can be found in Table I, with the difference in the definition of the symbols explained in the annotation.

### C. Resonant tunneling and rate equations

As discussed in Sec. II A, under the framework of tightbinding theory, the whole structure separated by the injection barrier is reckoned as a module containing the localized basis states. The electron transport across the injection barriers is modeled by resonant tunneling. The tunneling rate derived from the DM formalism is then used to describe the inter-module (interperiod) transport. As k-conservation holds for the first-order current, the tunneling rate from ji, ki to j j$^0$, ki is[15,50,71]

$$R_{ij_0,k} = \frac{2\Omega^2_{ij_0}\gamma_{ij_0,k}}{\Delta^2_{ij_0} + \gamma^2_{ij_0,k}}, \tag{8}$$

where j$^0$ denotes the subbands in the right-side module as illustrated in Fig. 2, $\gamma_{ij_0,k}$ is the dephasing rate, and $h\Delta_{ij_0} = E_{i,0} - E_{j_0,0}$ is the detuning energy from subband i to j$^0$. The dephasing rate consists of two parts, the lifetime broadening part from the intersubband scattering inside a module and the pure dephasing part between two subbands in different modules, reads[48]

$$\gamma_{ij_0,k} = \frac{1}{2}(\gamma_{i,k} + \gamma_{j,k}) + \gamma^{*}_{ij_0,k}: \tag{9}$$

The lifetime broadening is computed by summing over the inter-subband scattering rate to all subbands within a module $\gamma_{i,k} = \sum_{i' \neq i} W_{i',k}$, and the pure dephasing is calculated by accounting

the rate of IMP, IFR, and AD obtained through Eq. (7). The k-dependency needs to be removed for a 1D rate equation model. To do this, the first and second term in Eq. (9) is averaged before we compute the dephasing rate. The lifetime broadening can be calculated through the averaged scattering rate obtained by Eq. (4). The mean pure dephasing rate is resolved over the population difference between two subbands,[72]

$$\gamma^* \tag{10}$$

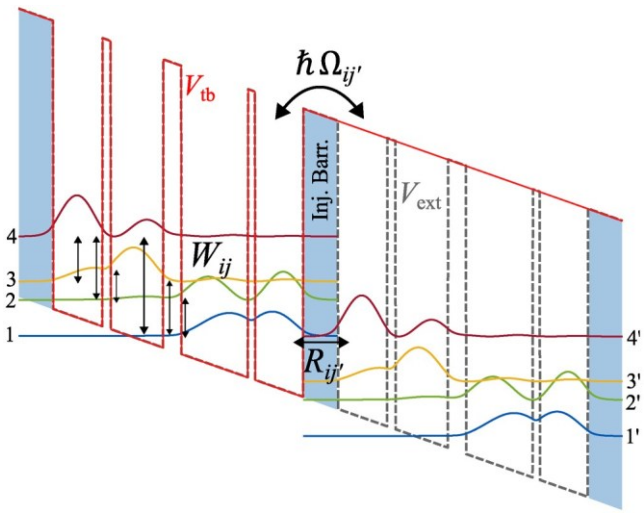



FIG. 2. Carrier transport model description. Two adjacent modules of a QCL are separated by the injection barrier (blue shaded area), with the tight-binding potential of the left period (solid red line) and extended potential (dashed grey) reported. Intra-module carrier transport is driven by inter-subband scattering. The coupling of two adjacent modules is described using coupling strength. The inter-module tunneling rate can be calculated with the detuning energy and the dephasing rate. The structure is from Ref. 70.

$$2\Omega_{2ij_0}\gamma_{ij_0}$$

$$R_{ij_0} = \frac{}{\Delta_{2ij_0} + \gamma_{2ij_0}}: \tag{12}$$

One could also obtain the mean tunneling rate by directly substituting Eq. (8) into (4), which we found nearly identical results in the L–I–V curve with the former way. However, all the results shown in this paper are obtained by Eq. (12) to avoid theoretical ambiguity. The last undefined parameter in Eq. (12) is the coupling strength (the Rabi frequency). Here, the coupling strength is calculated as[15,46]

$$(\hbar\Omega_{ij_0})^2 = D\,|\psi_i|V_{ext} - V_{tb}|\psi_{j_0}E D|\psi_i|V_{ext} - V_{0tb}|\psi_{j_0}E, \tag{13}$$

$$\gamma^*_{ij_0} = \frac{\pi\sum_{ij_0,fk_i j}(k_f - k_i)(k) - f_{j_0}}{}$$

$$(fk_{j_0})(jkkdk)jkdk : \text{Đ}$$

The dephasing rate can then be obtained as

$$\gamma_{ij_0} = (\gamma_i + \gamma_j)/2 + \gamma^*_{ij_0}: \tag{11}$$

The dephasing rate is then a k-independent value entering Eq. (8), and the k-independent first-order tunneling rate between doublets spanning the injection barrier is established with

where $V_{ext}$ is the extended potential profile for the complete two periods. $V_{tb}$ and $V^0_{tb}$ are the padded tight-binding potential for the left and right period respectively as shown in Fig. 2. Alternative ways can be used to calculate the coupling strength, for example, in Refs. 47 and 50. These two methods reported similar results to those from Eq. (13). A comparison of the coupling strength vs the electric field of the structure in Fig. 2 with all three methods is presented in Fig. 3. From Fig. 3 and also discussed in Ref. 47, Terazzi's method underestimated the coupling strength by about 20%, and because Razavipour's method sometimes reports complex solutions,[50] these two methods are not used in our model. Equation (13) will also underestimate the coupling strength by numerically comparing the calculated coupling strength of two ground localized bases with the detuning energy (anticrossing in actual QCL structure) $h\Delta$ of two identical quantum wells computed by the well-defined potential (extended) profile all at once. In Fig. 4, with the exact value of the coupling strength given by $h\Omega_{ect} = h\Delta/2$, we compute the ratio between the value from Eq. (13) and the exact value by varying the barrier width, well width, and the barrier height (i.e., the aluminum content in AlGaAs barrier). The ratio is generally stable when the barrier width is larger than 2 nm and slightly decreases with the well width. A lower ratio is also found for higher Al content. Generally, though some instability is found for thin well and barrier region, the approximation from Eq. (13) is accurate within the range shown in Fig. 4, with the worst scenario predicting 79% of the expected coupling strength at 15 nm well width and 0.5 nm barrier width with 0.3 Al content, indicating the overall validity of such method to compute the coupling strength for most of QCL structures. The method is also compared with the recent work, where the coupling strength is obtained by the EZ states (energy E and position z within the subspace of the multiplet) extracted from NEGF simualtion.[73] The coupling strength of the injector state and the upper lasing lever of the devices, ETH2019,[74] MITG552, MITG652,[75] and lU2022,[73] are investigated. The values in the same sequence with the above-mentioned device are calculated to be $\Omega$ = 1.37, 1.22, 1.32, and 1.67 meV. Reasonable agreement, with only 10% overestimation except ETH2019, has been reached by comparing with those from EZ states in Ref. 73, which are $\Omega$ = 1.58, 1.08, 1.17, and 1.49 meV. It is interesting to mention that when nonparabolicity is neglected, a better overall match is found (except ETH2019), and the results from our model is $\Omega$ = 1.22, 1.05, 1.12, and 1.44 meV. The reason for larger coupling strength is that when nonparabolicity is considered, the smaller effective mass in the barrier results in a larger overlap between the wavefunctions.

To this point, all necessary parameters used to establish the rate equations are given. Despite coherent and incoherent processes existing universally in the whole QCL structure, the states within a module are strongly coupled, and the doublets across the thick injection barrier are weakly coupled and strongly damped by dephasing. Hence, it is reasonable to treat the intra-module and inter-module transport by inter-subband scattering (incoherent) and resonant tunneling (coherent) respectively, as shown in Fig. 2. The rate equation of the two-period system can then be established. In principle, three periods are required in order to set up the rate equations, but due to the periodicity, only two periods are needed to close the quantum system. The population density of the state i in the primary period (left period in Fig. 2) is

$$\frac{dn_{si}}{dt} = \sum_{j,j=i} X\,(W_{ji} + R_{joi} + R_{jio})\,n_{sj} - n_{is} \sum_{j,j=i} X\,(W_{ij} + R_{ijo} + R_{ioj}): \quad (14)$$

In the above equation, the subscript $i^0j$ denotes the rate from the right period to the left period and vice versa. The $j = 1, \ldots, N$, where N is the number of subbands in one period. Note that the summation of j operate also on $j^0$ (e.g., in $R_{j^02}$, $_{j,P_{j=2}}$

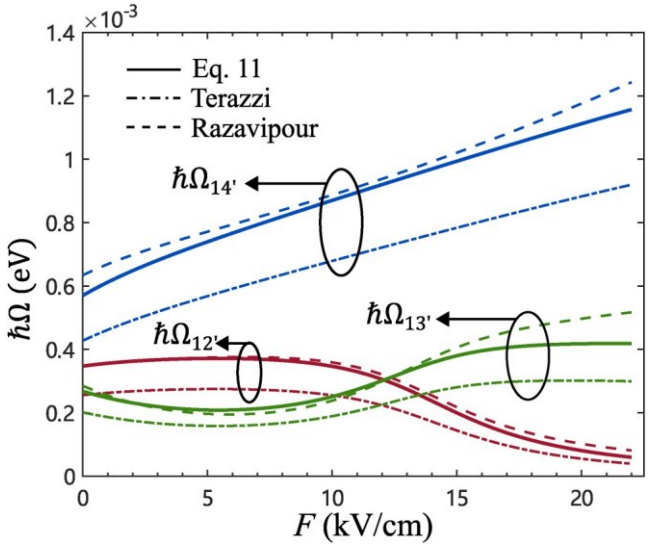
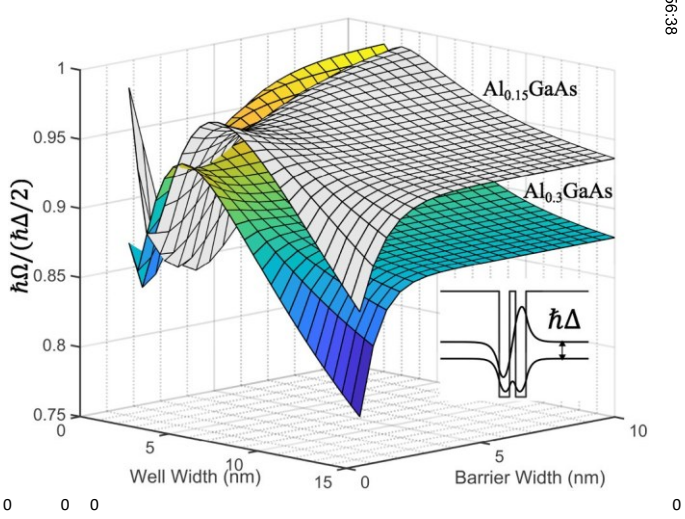



FIG. 3. Comparison of the coupling strength calculated by Eq. (13) (solid), Terazzi's method[47] (dashed-dotted), and Razavipour's method[50] (dashed) between states in the structure in Fig. 2 vs the electric field.

FIG. 4. The ratio between the calculated coupling strength and the exact value. The calculated value is obtained between the localized ground basis of two coupled QWs under tight-binding Hamiltonian from Eq. (13). The exact result is half of the detuning energy between the coupled states obtained by extended well-defined Hamiltonian as illustrated in the inset. The grey surface is for the $Al_{0.15}GaAs$ barrier, and the colored surface is for the $Al_{0.3}GaAs$ barrier.

$j = 1, 3, \ldots, N$). In Fig. 2, only four subbands are presented. Actually, there are two higher subbands not shown here as they have a negligible impact on carrier transport at low temperatures (less than 1% of the total carrier density in these two subbands at a lattice temperature of 77 K). In this scenario, the major subbands that participate in the carrier transport process are 1–4, and it is more computationally efficient to only include these subbands. It is worth noting that, for high-temperature operation, electrons can get enough kinetic energy to occupy higher subbands, resulting in additional current leakage. Thus, for high temperatures, all subbands should be included to study the electron transport behavior. Moreover, the theory describing the bound-to-continuum leakage (e.g., in Ref. 52) may also be used to better estimate the device performance at high temperatures. The steady-state solution of Eq. (14) can be obtained by setting $dn^s_i/dt = 0$. Because the scattering and tunneling rate are all related to the carrier density of each subband, the solution should be obtained in a self-consistent manner.[33] By assuming equal carrier distribution in all subbands initially, the new carrier densities of each subband in the next iteration can be calculated by Eq. (14) and renormalized according to the total sheet carrier density in one period, i.e., $\sum_i n^s_i = n^s_p$. The stability and convergence speed are improved by introducing a weighting factor dynamically during iterations.[34]




Experiment evidence has found that the electron temperature can be much higher than the lattice temperature.[76] The kinetic balance method is vital in the rate equation model to estimate the electron temperature that enters the Fermi–Dirac function. The latter describes the in-plane carrier distribution and is used to calculate the mean scattering rate in Eq. (4). The electron temperature will have a notable effect, especially on the scattering

rate with a cut-off wavevector, as elevated electron temperature will enable the hot electrons to have enough kinetic energy to scatter, which significantly increase the mean scattering rate between two subbands. The electron temperature is obtained by the kinetic energy balance method.[32,47,77] The kinetic energy generation rate reads

$$\delta K = \sum_{i,j} \sum_{m} n_{si} W_{ij(m)} h\Delta_{ij} + E_{o(m)}, \quad (15)$$

in which i, j sum over all subbands considered in one module, and m sums over all scattering mechanisms. $E_o^{(m)} = E_{ph}$ for LO phonon absorption and $E_o^{(m)} = E_{ph}$ for LO phonon emission, and $E_{ph}$ is the phonon energy. For the elastic process, $E_o^{(m)} = 0$. The tunneling rate is not included in Eq. (15) because we use first-order current here, where k-conservation holds, and the kinetic energy of the carrier does not change. If the second-order current is used, energy is conserved, and the resonant tunneling is an "elastic-like" process, and the tunneling rate must be added to Eq. (15).[47] Notably, experimental results have shown that the electron temperatures of certain structures can be different in each subband.[78] An enhanced kinetic energy balancing method, which determines individual $T_e$ per subband, was developed where a good agreement have been reached with experiment.[79] However, such a method could sometimes cause instability to the model because of the difficulties in solving sets of nonlinear equations. Some known problems have been encountered and discussed in, e.g., Refs. 47 and 80, including false solution at localized minimum, and overflow electron temperatures in certain subbands. Hence, to keep numerical robustness, we retain the single electron temperature model. Comparison with the Monte Carlo model in Ref. 15 shows the validity of our model, in which the electron temperature range extracted from the Monte Carlo simulation for the two-well structure at 10 and 125 K lattice temperature are 116–127 and 154–165 K, respectively, and the results from our rate equation model are 115.6 and 135.2 K, where the differences are all in a reasonable range.

The current density can be evaluated at the injection barrier after obtaining the electron temperature with a converged steady state. The equation for the current density reads

$$J = q_e \sum_{i,j} (R_{ij0} n_{si} - R_{j0i} n_{sj}), \quad (16)$$

where i sums over the states in the life period and j sums over the states in the right period.

### D. Carrier–light coupling

Starting from the optical properties of the QCL, the spectral gain coefficient can be calculated by[45]

$$g(h\omega) = \frac{q^2 e \omega}{\quad} \int |Z_{ij}|^2 \delta[f_i(k) - f_j(k)] L(h\omega - E_{ij,k}, \Gamma_{ij,k}) dk, \quad (17)$$

with the photon transition energy $h\omega$, the light speed c, the reflective index $n_r$, and the length of a period $L_p$. The summation is over all possible combinations of the states within one period. Diagonal gain between different periods is not considered here as the wavefunction is not well-defined in a tight-binding framework. $Z_{ij} = \psi_i | z | \psi_j$ is the dipole element. In addition,

$$L(x, \Gamma_h) = \frac{1}{\pi} \frac{\Gamma_h}{x^2 + \Gamma_h^2} \quad (18)$$

is the Lorentzian line shape function with half-width at half maximum (HWHM) $\Gamma_h$. $E_{ij,k} = E_{i,k} - E_{j,k}$ is the energy difference of subband i and j at wavevector k. $\Gamma_{ij,k} = h\gamma_{ij,k}$ is the k-dependent energy broadening. Equation (17) is a general equation for considering different in-plane nonparabolicity and k-dependent broadening. It can be simplified to the following equation when constant in-plane effective mass is used, and the broadening is averaged by Eqs. (4) and (10): $g(h\omega) = \frac{\quad}{\quad} \epsilon q_{02e} n \pi \omega r L_p \int_{E_i, E_j} |Z_{ij}|^2 (n_{is} - n_{js}) L(h\omega - E_{ij}, \Gamma_{ij})$: (19) c

From Ando's theory, the k-independent energy broadening $\Gamma_{ij} = h\gamma_{ij,k}$ can also be calculated using Eq. (11) by replacing $j^0$ with j,[15] i.e., considering the pure dephasing between subbands within a module. Different from the current broadening, when calculating the optical broadening, the lifetime part also needs to include the tunneling rate to adjacent periods $\gamma_i = \sum_{i' \neq i} (W_{i'} + (R_{i'0} + R_{i0'})$.

Without optical coupling, the result obtained from Eq. (19) is the unsaturated gain. The threshold current can be obtained when the maximum modal gain $\Gamma^{opt} g_{max}(h\omega_M)$ is larger than the modal loss

is the mode frequency, andαM, where ΓM is the optical confinement factor of modeαM is the total cavity loss. However, theM, ωM lasing field will also affect the carrier transport, altering subband populations and current density. Hence, it is important to include the carrier–light coupling after the threshold. We, therefore, introduce the optical transition rate between subband i and j,[54]

$$W_{ijopt} = \frac{}{M} c\epsilon q^{02} e n \pi r h \, jZ_{ijj2} \, X \, I_M L(h\omega_M \; E_{ij}, \Gamma_{ij}): \quad (20)$$

[53]The intensity of mode M, IM evolves over time domain with

$$\frac{@}{r}@I_M t = nc \, [\Gamma_{opt}{}^M g(h\omega_M) \; \alpha_M] I_M, \quad (21)$$

which can be further written in the intensity evolution over a short time interval Δt by[54]

$$I_M(t + \Delta t) = I_M(t) e^{[\Gamma_{Mopt} g(h\omega_M) \alpha_M]_{ngr} \Delta t}: \quad (22)$$

To reduce the computational load, mode competition can be disregarded by assuming single mode operation at maximum gain with photon transition energy Eul = Eu El from the upper lasing level (ULL) to the lower lasing level (LLL).[54] Thus, the summation in Eq. (20) is removed, and $h^{\omega}{}_M$ is replaced by Eul. Though $W_{ij}{}^{opt}$ here is a k-independent value, we can still substitute it into Eq. (4) to include the final state blocking effect. Carrier–light coupling can then be included in the rate equation model by adding the optical transition term $P (W^{opt}{}_{ji} n^s_j \; W_{ij}{}^{opt} n_i{}^s)$ in the right-hand side of Eq. (14).

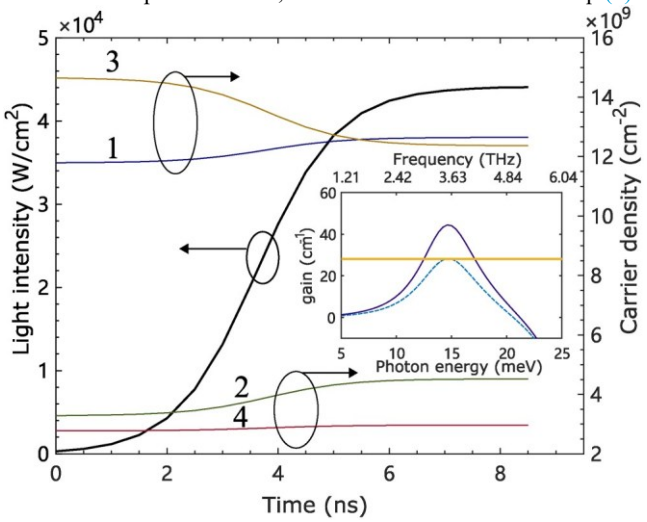



FIG. 5. In the four-well phonon–photon–phonon design shown in Fig. 2. Temporal evolution of the carrier density (right axis) and the light intensity (left axis) of the optical mode with the transition energy E32 = 14:3 meV (3.46 THz). The inset is the unsaturated gain and the saturated gain spectrum. In this simulation Te = TL = 77 K, cavity loss is set to be 28 cm$^{-1}$, optical confinement factor Γopt = 1, and interface roughness parameter are ΔIFR = 1:6 Å ΛIFR = 100 Å.

The iteration procedure of the carrier–light coupling is carried out as follows. First, the rate equation without the optical term is self-consistently resolved. The unsaturated modal gain can then be compared with the total cavity loss. Onceoptical coupling iteration begins. Since Eq. Γ(21)$^{opt}$Mgdoes not explicitly(hωM) . αM, the include the spontaneous emission term, a seed initial intensity [e.g., IM(t = 0) = 300 W/cm$^2$[54]] is applied to give the origin of the lasing oscillation. With a sufficiently short time interval (e.g., 0.3 ns), the optical transition rates from Eq. (20) can be calculated according to the present light intensity. Then, by assuming the optical transition rates to be constant within the short time interval, the rate equation containing the optical transition rate is selfconsistently solved, after which the subband population and gain are updated. The latter is then used to calculate the light intensity in the next time interval. The above procedure is repeated until a steady state is reached for IM. Note here that the cavity loss consists of waveguide loss and mirror loss, and the optical confinement factor can be obtained by waveguide modeling techniques or experimentally result. These values are treated as predefined input parameters in this model.

Figure 5 shows the simulated temporal evolution of the fourwell structure in Ref. 70 (the same structure in Fig. 2). During the carrier–light coupling evolution, the carrier density of state 3 decreased. While the carrier density of state 2 increased because of the increased optical transition rate. Steady state is reached at ~8 ns. The saturated gain at a steady state compared with an unsaturated gain is plotted in the inset of Fig. 5, and the peak of the saturated gain coincides with the cavity loss (28 cm$^{-1}$).

## E. Model layout

The complete flow chart of the model in this study is shown in Fig. 12 in the Appendix B. As can be seen, the self-consistent procedure of the rate equation, together with the carrier–light coupling, is nested in the kinetic balancing loop. The form factors of IMP, LO phonon, and EE scattering are calculated before the kinetic balance loop is conducted. Next, the k-dependent scattering and dephasing rates are calculated, and form factors are visited according to the specific exchange wavevector with the interpolation method. As mentioned in Sec. II B, the Debye and Thomas– Fermi screening method is used. The k-dependent scattering rates, dephasing rates,

and coupling strength do not change in the kinetic balance loop. Thus, multiple invoking these time-consuming calculation subroutines are avoided, and scattering between subbands only needs to be averaged using the existing data by the Fermi– Dirac distribution. The outermost Schrödinger–Poisson loop is for the "self-self-consistent" solution. As the electrostatic potential alters the potential profile, wavefunction and eigenstates are changed. Hence, it is inevitable to recalculate form factors and the k-dependent scattering rate in each Schrödinger–Poisson iteration. The total computational time is multiplied by the convergence iterations needed for the Schrödinger–Poisson loop. A possible way to avoid this heavy computational load is to solve the Schrödinger– Poisson system before the transport model, i.e., from thermal distribution. However, significant differences in the potential profile and gain spectra have been found between these two
methods.[47,81]

The computational efficiency is greatly improved as we calculate the look-up table of form factors before commencing the multi-dimensional integration of scattering mechanisms like IMP and EE. Because the k-dependent scattering rates in the mesh grid of wavevector space are pre-calculated, they only need to be averaged within the self-consistent procedure. In a modern PC with an Intel i9 13900K processor, and considering a four-level structure, with around 1000 and 50 discrete points in real and k space respectively, the MATLAB (MathWorks Inc.) implementation of the full calculation for a single bias typically takes about 2.5 min including the carrier light coupling with Schrödinger–Poisson equation converged in five iterations. Fast variation of the structural parameters can be achieved by ignoring EE scattering (using about 60% of the total time) and compensating by IFR scattering, as suggested in Ref. 82. The Schrödinger–Poisson system can also be disregarded for low doping densities. With these two simplifications, an L–I–V curve with 60 points can be calculated within 10 min, moving the code to a server with Intel Xeon Platinum 8468 only half the computational time. Such small improvements could be attributed to the poor multiplicity of the code, where many sequential procedures still exist in the code. Potential acceleration of the efficiency can be done by optimizing the algorithm, e.g., calculating the form factor using discretized Fourier transform,[51,83] and multiplicity or moving the code to some low-level programming language such as C++ and FORTRAN. This is, however, beyond the topic of this paper, so we will not do further discussion here, but it is obvious this configuration of the program has much greater potential to be used as a fast optimization tool of QCLs.

## III. RESULTS AND DISCUSSION A. L–I–Vcharacteristics

We use three devices ranging from two- to four-well design to verify the model and compare the calculated L–I–V curves with experiment results. The band diagram and their wave functions are given in Fig. 6. The first device (EV1183)[84] is a two-well design in which the population inversion is established between j3⁰i and j2⁰i. The electrons are injected from j1i to j3⁰i by resonant tunneling and carrier depopulation is achieved by LO phonon scattering from j2⁰i to j1⁰i. The center 15 Å of the phonon well is doped to give a 1:5 $10^{10}$ $cm^2$ sheet carrier density per period. This is a typical design of a three-level system. Note that other higher bound states also exist and may involve in the carrier transport at high temperatures, resulting in additional leakage channels that sabotage the population inversion. It is worth mentioning that the record of operation temperature so far was achieved by the two-well design.[5,75] A clean three-level system is achieved by using a higher barrier and reducing the leakage parasitic channels. The latter is found to be crucial for high-temperature operation of QCLs. The second device (V775) is the one with operation temperature up to 200 K.[62] It has three wells, and the center 50 Å of the widest phonon well is doped to give 3 $10^{10}$ $cm^2$ sheet carrier density per period. The carrier transport behavior is very similar to the two-well design. The difference is that an additional subband j2⁰i exists near j3₀i, population inversion is created between2₀i. Because of the broadening of the optical line-j4₀i and both j3₀i and j
width, both could contribute to the gain of the QCL. Additionally, the 41 Å wide barrier to the left of the phonon well is almost comparable with the injection well (43 Å). A probably more accurate way of modeling this structure could be reckoning this barrier as an additional coupling barrier under tight-binding theory, and thus the transport from j3i to j2i is described by resonant tunneling. This, however, has not yet been included in the current model. The third device (V843) with four wells has two LO phonon scattering processes. The injection barrier is delta doped to give a 3:25 $10_{10}$ $cm_2$ sheet carrier density per period. Carriers injected from j1i to j4₀i followed by an LO phonon scattering from j4₀i to j3₀i. The electron depopulation is again achieved by LO phonon relaxa-
$3_0$
tion frombetweenThe calculatedj ij2and⁰i toj$2_0$ji1L.⁰–i. Population inversion is, thus, establishedI–V curves under different temperatures are

compared to the experimental measurements and other models in Fig. 7. The material parameters of AlGaAs are calculated by Vegard's law with the data from Ref. 87. Because all three devices used a double metal waveguide, the optical confinement factors are set to be 1 in the simulation. The total cavity losses are set according to the original study of these devices. They are 12 $cm^{-1}$ for device EV1183,[84] 37.5 $cm^{-1}$ for V775,[62] and 38.2 $cm^{-1}$ for V843.[70] In the characteristic conducted in this section, the interface roughness correlation length $\Lambda_{IFR}$ is all set to be 100 Å for these structures. The mean height $\Delta_{IFR}$ for EV1183 and EV775 is 2 Å. It is, however, set to 1.2 Å for V843 to better fit the experimental result. We also assume the lattice temperature is the same as the heatsink temperature in pulsed mode

reported in the original studies. The light intensities are renormalized to present a clear comparison with the experimental results.

For the two-well resonant phonon device (EV1183), the calculated I–V curve at 10 K generally matches the experiment data except for an additional current peak at 9 kV/cm. This bump is attributed to the alignment of state j1i and j2oi, where the tunneling rate $R_{12}^{o}$ at 9 kV/cm is calculated to be $1.35 \times 10^{11}$ s$^{-1}$. The experimental measurement result does not show any current peak or a plateau. The discrepancy is partially due to the growth deviation,[15] which causes well width to vary in different stages. Additional broadening of the current density exists as the alignments are reached at different biases across the whole structure, which may flatten the current peak. Our calculation perfectly repeats the I–V curve calculated using the density matrix ensemble Monte Carlo method (DM-EMC),[15] where the current peak is also found at the same bias. We also show the L–I–V curve under different temperatures. The simulated L–I–V curve does not change significantly with temperature from 10 to 120 K. The bent I–L curve at 10 K near the threshold is due to the lasing starting at the negative differential resistance (NDR) region. The current after this threshold bias starts to increase because the optical current appears, and resonance condition is gradually established between state j1i and j3oi. The current reaches a second peak (0.77 kA/cm$^2$) when j1i is aligned with j3oi at 13.5 kV/cm. As mentioned before, our model considers only the mode with the highest gain, while in reality, multi-mode behavior and mode hopping effects may exist.[82] This may explain why the calculated I–L curve has a generally

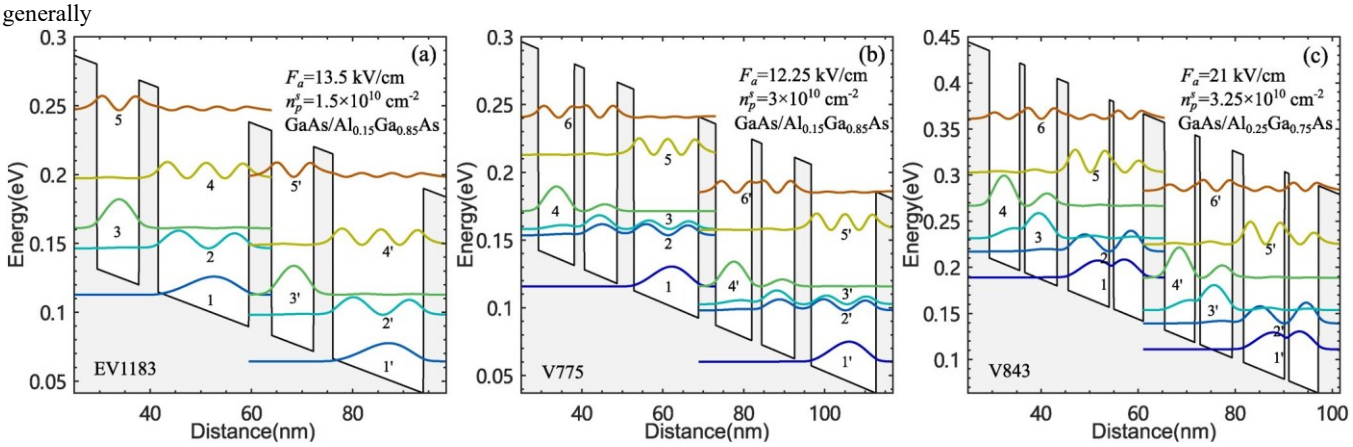



FIG. 6. Conduction band potential profile and probability densities offset by the eigenenergies of the investigated structures (two periods). Bias, sheet carrier density per period, material composition, and wafer number are reported. (a) Two-well resonant phonon device (EV1183).[84] (b) Three-well resonant phonon device (V775).[62] (c) Four-well indirect pumped phonon–photon–phonon device (V843).[70]

linear dependency while the experimental one is bell-shaped in this structure or noisy in other devices. The threshold currents $J_{th}$ at 10 and 70 K are 0.3 and 0.36 kA/cm$^2$, respectively, which match the experimental results. However, the model fails to estimate the threshold current at 120 K. The experiment finds a very steep increase in threshold current after 70 K. At 120 K, the experimental threshold is around 0.7 kA/cm$^2$, but the model obtains $J_{th}$ ¼ 0:32 kA/cm$^2$, which only about half of the experimental result. Although the waveguide loss could increase with temperature,[88] changing the waveguide loss at the same temperature in the model cannot reproduce such a result because the simulated maximum of the non-lasing I–V curve is only 0.38 kA/cm$^2$ at alignment bias with such temperature. The model, however, predicts 0.7 kA/cm$^2$ non-lasing peak current at a much higher operation temperature of 280 K. From an experimental point of view, the operation temperature can be higher than the heatsink, even with pulsed operation.[11] Still, it seems this effect cannot compensate for such a significant difference. This may be due to the thermally activated strong current leakage over the barriers to the continuum at high temperature as the barrier of this device is relatively low (15% Al content in the barrier). Hence, this inconsistency with the experiment can be potentially improved by adding a proper thermal model and considering the current leakage to the continuum.

Our calculation shows an excellent agreement with the experiment for the second resonant phonon device (V775). At 10 K, the small bump at 2.5 kV/cm is due to the alignment of j2i and j3oi. Notably, the calculated I–V at 10 K peaks at 9 kV/cm, followed by an NDR region before the threshold (9–10.25 kV/cm). The peak is to drop beyond this alignment and again increases as the designedattributed to the alignment between j1i and j3$^0$i. The current starts injection channel between j1i and j4oi is turned on, which corresponds to the third current peak at 12.25 kV/cm. The calculated NDR region coincides with the experimental current plateau at the same bias range. The reason why the experiment shows a plateau instead of a valley-shaped curve has been explained by the electric field domain (EFD) in Ref. 85. In short, because the device tends to maintain current flow continuity, another current carrying channel with a higher electric field starts to establish from the top contact side of the active region. The total applied voltage increased as more and more periods switched to the higher EFD, but the current almost remained the same. This hypothesis was later proved in Ref. 86 by using scanning voltage microscopy (SVM) with the device (V843) that we will discuss in the following paragraph. Apart from the I–V, the I–L curve fits quite well with a waveguide loss of 37.5 cm$^{-1}$. The underestimated threshold at low temperatures may also be attributed to the EFD effect. At higher temperatures, from 80 to 150 K, the simulated threshold matches well with the experimental one. However, with the increased temperature, an additional simulated current plateau appeared from 7 to 9 kV/cm because the resonant tunneling fromj j1i 1toj 2oi with a lower bias is equally important as tunneling from j to j3oi with a higher bias. A generally more stable transport property is found, and the threshold for higher temperature is located at the positive differential resistance (PDR) region, which makes it easier for the model to estimate the threshold current. Last, we find the simulated lasing I–V curves are almost identical for different temperatures with different non-lasing curves pinned on them. This behavior also matches the measurements at different temperatures in Ref. 85.

In the simulation of the four-well phonon–photon–phonon device (V843), a smaller interface roughness mean height is chosen $\Delta_{IFR}$ = 1.2 Å to fit the measurement result. An overall agreement is achieved with the experiment. The result from the rate equation method in this work is comparable with the NEGF result.[70] A lower peak current is found at the designed bias (21.5 kV/cm) in the NEGF

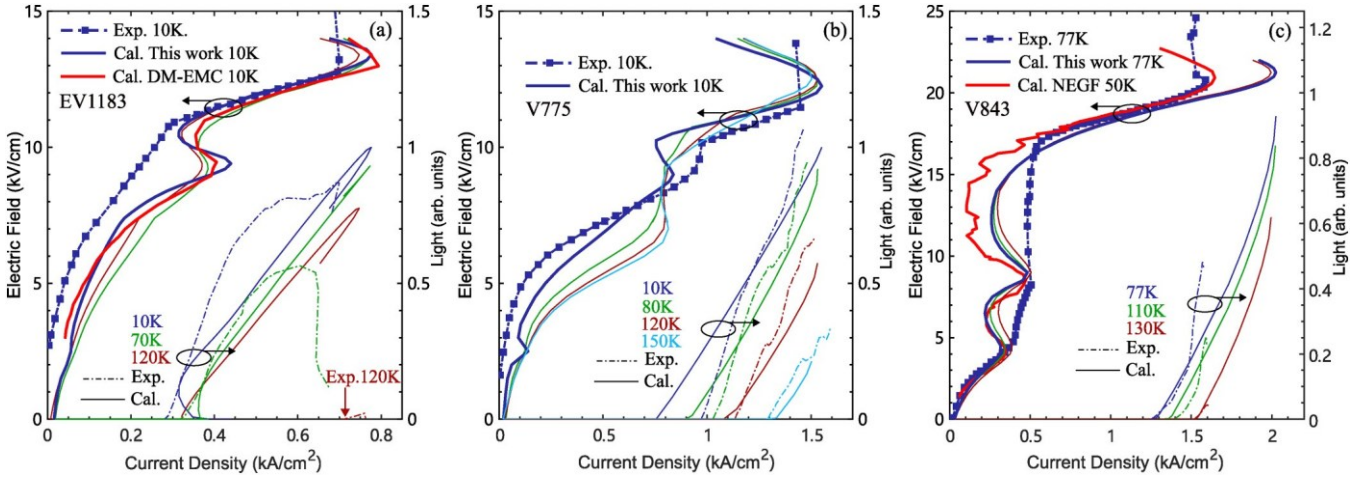



FIG. 7. L–I–V curves of the three structures from the model in this work, experiments, and other models. (a) EV1183: experiment data are extracted from Ref. 84, and DM-EMC result is from Ref. 15. (b) V775: experiment measurement is from Refs. 62 and 85. (c) V843: experiment data from Ref. 86, and NEGF calculated in Ref. 70.

simulation because carrier light coupling was not considered in the NEGF method.[70] NEGF and our model predict two current peaks before the threshold at 4.5 and 9 kV/cm. The former is because of the tunneling from j1i to j2⁰i, the latter is because of the alignment of j1i and j3$_0$i. The difference between the experimental I–V plateau from 9 kV/cm to the turning point before the current sharply increases and the valley-shaped simulation result is again because of the EFD. SVM measurement has shown that in the current plateau from 9 to 16 kV/cm, two EFDs coexist and are pinned at these two ends of electric fields.[86] Reproduction of this phenomenon may need a deeper understanding of the dynamics of the EFD formation,[89] which could be interesting to be included in the models in the future version. Note that in another current valley from 4.5 and 9 kV/cm, where a current shoulder is found in the experiment, a uniformed electric field is revealed by SVM across the whole structure in this region.[86] Hence, it is likely the difference between the theoretical and experiment here is not due to the EFD but because some unknown broadening mechanisms fade the valley away from the I–V curve. Our model predicts the correct threshold current at different temperatures but with a larger dynamic range. The theoretical peak current is around 2 kA/cm$^2$ at 21.25 kV/cm comparable to 1.6 kA/cm$^2$ at 20.8 kV/cm in the measurement. The model overestimates the dynamic range because, experimentally, the dynamic range can be cut off because the driving circuit pushes the laser into the following NDR region before it reaches the designed alignment bias. This phenomenon has also been found in the V775, as the same design with different metal contact shows a significant difference in dynamic range.[62]

From the above-mentioned comparison, our model reproduces the experiment measurements with reasonable accuracy. Difference between the calculation and experiments still exists because many complicated phenomena and effects from reality are not fully revealed in the model. Nevertheless, a perfect agreement with the results from counterpart models, DM-EMC and NEGF, shows the rationality of the difference with the experiment and further validates our model. Fast computational speed enables us to contain carrier light coupling and self-self-consistent Schrödinger– Poisson system with reasonable calculation time. As THz QCL designs are numerically and experimentally very sensitive, further improvements in the accuracy may require more realistic effects to be considered in the model, which may inevitably bring about a heavier numerical load. By intensively testing different designs, the current model seems adequate to fulfill the requirements to predict the performance of a QCL quantitatively. Furthermore, the microscopical information of the active region can help to identify some optimization directions, such as the influence of the barrier height and doping, trade-off between diagonal and vertical transition design, suppressing the leakage channel, etc., showing the broad application scenario of the model.

### B. Influence of the interface roughness parameter

The interface roughness is vital to the QCL performance. As it is hard to directly measure the local imperfections,[90] it is treated as two effective parameters, correlation height $\Delta_{IFR}$ and length $\Lambda_{IFR}$, describing the average deviation in the heterostructure. These two parameters are usually swept in the range of $\Delta_{IFR}$ ¼ 0:1–3 Å, $\Lambda_{IFR}$ ¼ 20–100 Å.[82] This section analyzes the influence of varying IFR parameters on our rate equation model. IFR is considered in the inter-subband scattering rate, inter-module pure dephasing rate, and intra-module pure dephasing rate in the model. These three aspects further affect the subband population, dephasing time, optical linewidth of the structure, and the general performance of the device. The degree of the impact of IFR depends on specific QCL designs. In the four-well design V843, with much thinner barriers within a period, nonzero wavefunctions considerably extended across the interfaces, and the influence of IFR parameters becomes more critical. Here, we use the V843 design to demonstrate the effect of the varying IFR parameters in this model. To reduce the complexity of this analysis, we fix the correlation length $\Lambda_{IFR}$ to a typical value of 100 Å,[53] varying the height

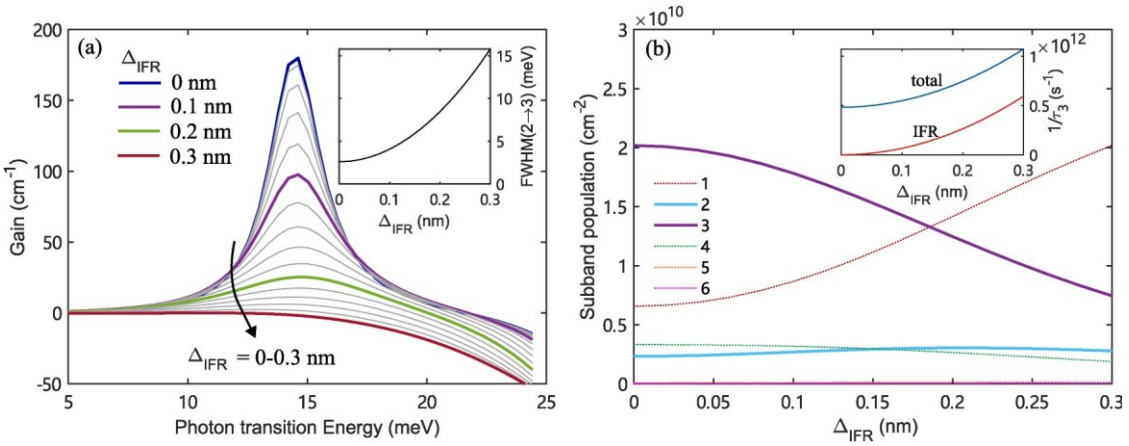



FIG. 8. Influence of the IFR parameter to the unsaturated gain of V843 at designed bias 21.5 kV/cm, $\Lambda_{IFR}$ is fixed to 10 nm, $T_L$ ¼ 77 K, $\Delta_{IFR}$ varies from 0 (no IFR) to 0.3 nm. (a) Gain spectrum with varying correlation height. Inset is the optical linewidth of transition 2 → 3 vs the $\Delta_{IFR}$ parameter. (b) Subband population change with the $\Delta_{IFR}$. Inset is the total inverse lifetime (blue) of the ULL j3i and contribution from IFR scattering (red).

parameter $\Delta_{IFR}$, and focus on the impact to the gain and current density. Note that carrier–light coupling is turned off in this test.

Figure 8 shows the gain evolution along with the IFR correlation height. Test conditions are given in the figure caption. At the designed alignment bias of V843, the gain peak is dramatically dropped when $\Delta_{IFR}$ increased from 0 to 0.3 nm. Although IFR is related to many properties of the QCL, such as carrier lifetime, tunneling rate, and electron temperature, two key factors contribute to such a change. First, a larger $\Delta_{IFR}$ corresponds to a larger IFR scattering rate and pure intra-module dephasing rate, leading to the increased optical linewidth, which is shown in the inset of Fig. 8(a). For the parameter used in Sec. III A $\Delta_{IFR}$ ¼ 0:12 nm, the calculated FWHM is about 5 meV (1.2 THz), which meets the typical optical linewidth for THz QCL.[7,50,51] Second, a larger $\Delta_{IFR}$ results in a higher inter-subband scattering rate, making the population inversion more difficult to be achieved. In Fig. 8(b), the subband population of the ULL j3i decreased from 2 10^10 to 0:75 10^10 cm^2 when $\Delta_{IFR}$ increased from 0 to 0.3 nm. With a higher IFR scattering rate, more and more carriers are scattered from j3i to j1i through indirect (3 → 2 → 1) and direct (3 → 1) channels. Such a change can also be reflected in the inversion lifetime of j3i, the increment of $1/\tau_3$ is apparent because of the contribution from IFR. It is worth mentioning that although the population inversion is still held when $\Delta_{IFR}$ ¼ 0:3 nm, no gain has been observed because the tail of the strong absorption peak at phonon energy (~36 meV) from j1i ! j2i and j3i ! j4i acts as material loss and overcomes the gain from j3i ! j2i at working frequency.

Apart from the gain spectrum, the change of the I–V curve with varying $\Delta_{IFR}$ is given in Fig. 9. The broadening effect on the I–V curve is pronounced. No local peak current is found when $\Delta_{IFR}$ = 0.3 nm at 4.5 or 9 kV/cm. Instead, a continuously increasing I–V is observed. It is known that the broadening of the I–V curve is very sensitive to the pure dephasing rate.[47,51] From the inset of Fig. 9, we can see that the pure dephasing rate is dominated by the IFR contribution. AD scattering plays a minor role in the GaAs/ AlGaAs material system, as most wavefunctions are localized in the well (GaAs) region. V843 is delta doped at the injection barrier, which is less overlapped with the wavefunctions, making the IMP scattering less dominant. Also, from the relation of the peak gain vs the electric field, the gain is observed and reaches its peak at a similar applied field, with the dropping magnitude for larger $\Delta_{IFR}$. Then, for the same cavity loss of 38 cm^-1, the threshold is found at a higher electric field and current density for a larger correlation height. Lasing cannot occur when $\Delta_{IFR}$ is as large as 0.2 nm in this

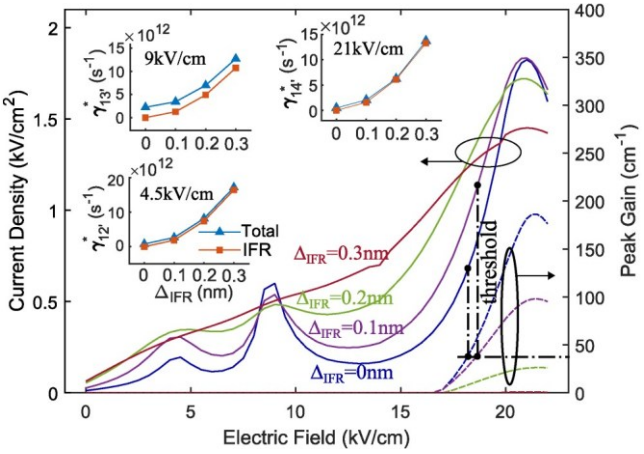



FIG. 9. I–V and peak optical gain vs the electric field of V843 for varying $\Delta_{IFR}$ from 0 to 0.3 nm, $\Lambda_{IFR}$ ¼ 10 nm The three insets are the total and IFR pure dephasing rate of 1–2^0, 1–3^0, and 1–4^0 at their alignment electric field of 4.5, 9, and 21 kV/cm, respectively.

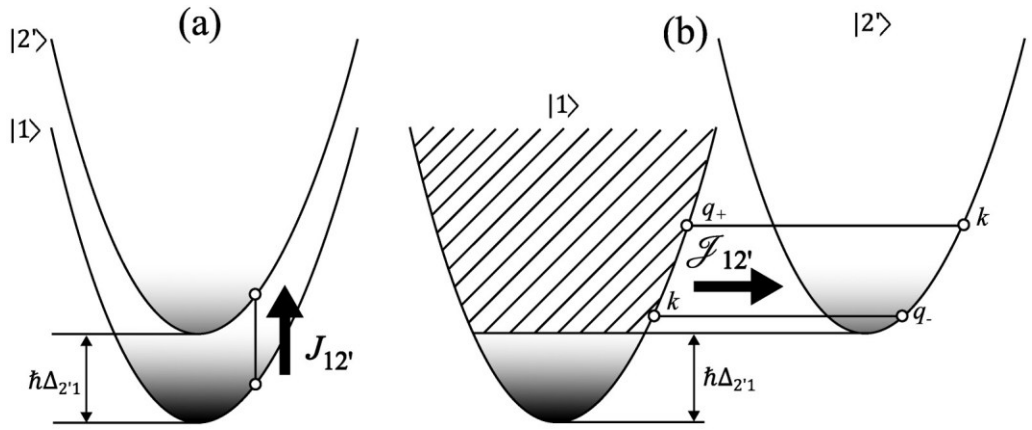



FIG. 10. Schematic diagram of the (a) first-order current approximation and (b) second-order current approximation.

test. This, to some extent, indicates that the growth quality is essential to the performance of QCLs.

From a modeling point of view, the reproduction of IFR in the actual device has always been troublesome because it can vary from hundreds of interfaces and different devices with even identical active region designs. It is also one of the major gaps that prevent a truly ab initio model, as the IFR parameters still need to be treated as a phenomenological fitting parameter. As suggested in Ref. 11, a larger IFR parameter can be used to represent yet unknown mechanisms and compensate for the EE scattering. A similar result can be achieved with larger IFR parameters, and the computational efficiency can be greatly improved without the EE mechanism.

### C. Second-order current

Although second-order current approximation is disregarded in the above test results, it may be helpful to clarify and discuss the influence and difference with the first-order current here. The second-order current is implemented by incorporating the firstorder tunneling rate with the correction parameter σ.[9,47,71] The second-order tunneling rate reads: $R_{ij_0} = R_{ij_0}\sigma_{ij_0}$, and $\sigma_{ij_0}$ is calculated by

$$\sigma_{ij_0} = \Theta(h\Delta_{ij_0}) + \Theta(h\Delta_{j_0i})\ln\{1\ln\{1^{+}\exp[(+\exp[(\mu_i - \mu E_i - E^{hi})\Delta/(jk_oiB)/T(ek)]\}_BT_e)]\}, \tag{23}$$

where $\mu_i$ is the Fermi level of subband i. Θ(x) is the Heaviside step function. The second-order current density is obtained by replacing $R_{ij_0}$ with $R_{ij_0}$ in Eq. (16). Its application in the rate equation model is equivalent to reckoning the tunneling rate as an elastic-like process.[9,47,71] As demonstrated in Fig. 10(a), the first-order current is established by the population difference of two states with identical wavevector k. Hence, no matter whether the final subband energy is higher or lower than the initial subband energy, all electrons in the initial subband will contribute to the current. The second-order current approximation shown in Fig. 10(b) follows the energy conservation. When the subband edge of the initial subband is lower than the final subband ($E_1$, $E_2$), only the electrons located in the shaded area, where $E_{1,k}$. $E_2$ will contribute to the total current from j1i to j2i.

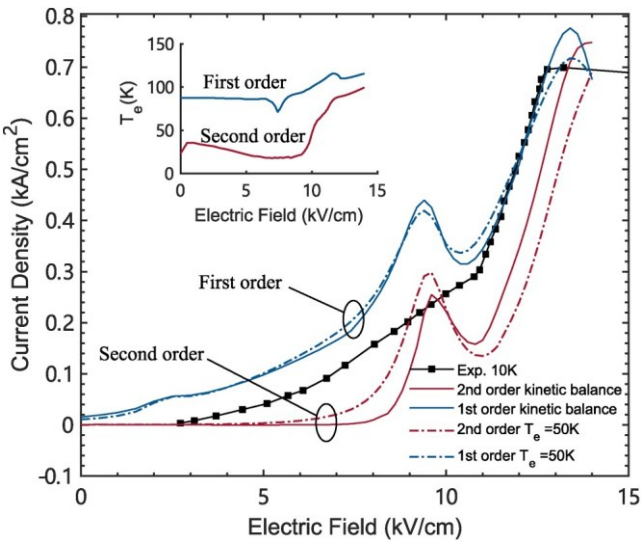



FIG. 11. I–V comparison between first-order and second-order current approximation of the device EV1183, with kinetic balance and by assuming a constant electron temperature of 50 K. The lattice temperature is set to be 10 K in this test. The inset is the electron temperature estimated by the two methods.

As an elastic-like process, the tunneling rate must be included in the kinetic energy equation if the second-order current approximation is applied in the model. Equation (15) becomes

$$\delta K = \sum_{i,j} \sum_{m} n_{si} W_{ij(m)} (h\Delta_{ij} + E_{o(m)}) + \sum_{i,j_0} h\Delta_{ij_0} (R_{ij_0} n_{si} - R_{j_0 i} n_{sj}). \tag{24}$$

The second term in the above equation accounts for the kinetic energy rate from resonant tunneling. With these two modifications, EV1183 is used to test the difference between the first and second current approximation. Figure 11 shows the I–V test result using both methods. The experiment result is also plotted in the same figure for side-by-side comparison. With the kinetic balance method, the current density is generally lower for the second-order approximation with a current peak at the same bias. Almost zero current density is observed until the electric field reaches 8 kV/cm. The second-order approximation also predicts lower electron temperature. Such difference is more prominent at the low applied field, which is, however, not because of the second term in Eq. (24) but due to a significantly smaller kinetic term from incoherent scattering [first term in Eq. (24)]. Because the second-order approximation largely suppresses the tunneling from a lower subband to a higher subband (i.e., j1i to $j2^0$i) before they aligned, the subband carrier densities at zero bias for j2i are 7:3 $\times 10^8$ cm$^2$ and 2:8 $\times 10^7$ cm$^2$ for first- and second-order approximation, respectively. With much higher carrier scattered from j2i to j1i in the first-order approximation, a higher electron temperature is found to represent these scattered electrons to the lower subband j1i. To eliminate the influence of the electron temperature, a second test is conducted with a constant $T_e = 50$ K and other input parameters unchanged. With constant $T_e$ (lower than the $T_e$ estimated by the kinetic balance method from the first-order model and higher than the second-order model at the low applied field from 0 to 9 kV/cm), the I–V for the first-order model is almost unchanged and current from the second-order model has slightly increased below 9 kV/cm. By comparing with the experiment result, we found that the second-order approximation significantly underestimates the current density with the kinetic balance method. Although an improved result was found by artificially setting a higher constant electron temperature, $T_e = 50$ K, such a difference can still not be compensated. From the definition of the second-order parameter σ, the difference with the first-order current becomes smaller for higher temperatures as the slope of σ vs $h\Delta_{j0i}$ becomes less steep. For low electron temperature, $\sigma(h\Delta_{j0i})$ becomes more abrupt and it decays to the step function $[\sigma_{ij0} = \Theta(h\Delta_{ij0})]$ as $T_e$ approaches to zero. For low temperatures, carriers are settled in bottom of j1i, and no current flow will be found. Apart from the device shown here, test results from other structures also show underestimated current density at low temperatures. This leads us to suspect that the second-order current approximation implemented in the rate equation model may underestimate the tunneling rate, especially at low temperatures (i.e., $k_B T_e \ll h\Delta$).[47] Derived from the DM, the coherence associated with the transition j2, ki to j1, ki consists of the direct contribution (first order) and scattering-assisted contribution (second order).[47,91] The current density is established between four states with two additional exchange wavevectors $q_+$ having same energy with the other subband at wavevector k [e.g., in Fig. 10(b), they are j1, ki, j1, $q_p$i, j2, ki, j2, qi, with $E_{1,k} = E_{2,q}$ and $E_{2,k} = E_{2,q_p}$]. By implementing the second-order current in the rate equation, contribution from j1, ki with lower energy than the subband edge of j2i (e.g., $E_{1,k}$ , $E_{2,0}$) is completely ignored, which is probably the reason of such underestimation at low temperatures before alignment is reached. Validation of the second-order current theory is beyond the limit of this study, which may need more theoretical and experimental effort. As many aspects can result in the difference between reality and theory, accurately modeling THz QCL remains an open topic. In this study, although it seems that the first-order current can better fit the experimental result at low temperatures, the model indeed overestimates I–V a bit at a low bias range, so it could still be important to consider the correction from the second-order current in our rate equation model, but in a more comprehensive way.

## IV. CONCLUSION

In this study, a rate equation transport model, including resonant tunneling from DM formalism, is developed. Theoretical and model description is given in detail, with some analysis of the intermediate parameters during the simulation. The model follows the spirit of ab initio modeling, and the only fitting parameter in the transport model is the interface roughness correlation height and length. The influence of the IFR parameters has been presented in Sec. III B. In Sec. III A, three devices with different design strategies have been modeled with good agreement with experiment measurement and other models, such as DM-EMC[15] and NEGF,[70] showing good compatibility of the model for various designs. Our model is highly computationally efficient and flexible compared to counterpart models. The calculation time for a complete simulation at a single bias is about one minute in a modern server with adequate numerical settings to maintain accuracy. Moreover, the Schrödinger–Poisson system, kinetic energy balancing, and carrier light coupling can be turned on and off according to specific application scenarios. For example, for fast variation, the self-self-consistent Schrödinger–Poisson equation can be disregarded, and the result of a single bias can be extracted in 10 s. Benefiting from the arbitrarily defined number of subbands entering the transport, current leakage to continuum in some cases may be compensated by including a higher quasi-continuum state from the FDM, despite the specific current leakage model not being included yet. A deep insight into the transport process can be revealed by various intermediate parameters, e.g., scattering rate, tunneling rate, dephasing time, unsaturated and saturated gain, etc., providing an intuitive understanding of complicated phenomena in QCLs. The flexibility and computational efficiency of the model enable us to optimize the structural

parameters with fast variation speed and sufficient information extracted from the results. The secondorder current approximation does not fit well the experiment result at low temperatures, whereas the first-order current model is more suitable in such conditions. This may indicate the necessity to develop a more sophisticated theory to implement such correction to the rate equation model. More realistic effects and improvements, including continuum current leakage, hot phonon, multi-mode carrier–light coupling, second-order gain, and tight binding with arbitrary coupling barrier, could be incorporated into the current model to push the limit of accuracy further—this will, of course, albeit at the expense of greater computational load. Overall, despite lots of simplification has been made, we conclude that the current configuration of the model has a good balance between computational speed and accuracy. Therefore, this model can be used as a fast design optimization tool and give a first insight into the design performance of THz QCL.

TABLE I. The source of the scattering rate equations used in this model and corresponding equation numbers in the reference.

| Scattering mechanism | Symbol in this work | Symbol in reference | Reference | Equation number in reference |
|---|---|---|---|---|
| LO[a] | $W_{ij(LO,k_i)}$ | $W_{ik,j}$ | 53 | Eqs. (82)–(87) |
| EE[a] | $W_{ii(EE0jj0),k_i}$ | $W_{ijfg(EE)}(k_i)$ | 51 | Eqs. (17)–(18) |
| IMP (inter-subband)[a] | $W_{ij(IMP,k_i)}$ | $W_{ik,j}$ | 53 | Eqs. (101)–(103) |
| IFR (inter-subband)[a] | $W_{ij(IFR,k_i)}$ | $W_{ik,j}$ | 53 | Eqs. (110) |
| AD (inter-subband)[a] | $W_{ij(AD,k_i)}$ | $W_{if}(k_i)$ | 32 | Eq. (10.248) |
| IMP (pure dephasing)[b] | $\gamma^*_{(ij,kIMP)}$ | $\Gamma_{(intra\mu,v)}$ | 47 | Eq. (5.54) |
| IFR (pure dephasing)[b] | $\gamma^*_{(ij,kIFR)}$ | $\Gamma_{(intra\mu,v)}$ | 47 | Eq. (5.27) |
| AD (pure dephasing)[b] | $\gamma^*_{(ij,kAD)}$ | $\Gamma_{(intra\mu,v)}$ | 47 | Eq. (5.45) |

[a] The physical definition of the symbol in this work is the same as the references.

[b] The relation of the physical definition between this work and the reference is: $\gamma_{0ij} = \Gamma_{intra(i,j)}/2\hbar$.


## ACKNOWLEDGMENTS

We acknowledge support from the Natural Science Foundation of Guangdong Province of China (Grant No. 2023A1515012793) for funding the research.


## AUTHOR DECLARATIONS Conflict of Interest

The authors have no conflicts to disclose.

## Author Contributions

Zhichao Chen: Data curation (lead); Methodology (lead); Software (lead); Writing – original draft (lead); Writing – review & editing (equal). Andong Liu: Data curation (supporting); Software (supporting). Dong Chang: Project administration (supporting); Resources (equal). Sukhdeep Dhillon: Investigation (equal); Writing – review & editing (equal). Manijeh Razeghi: Investigation (equal); Writing – review & editing (equal). Feihu Wang: Funding acquisition (lead); Investigation (lead); Supervision (lead); Writing – review & editing (equal).

## SUPPLEMENTARY MATERIAL

The source code and documentation for its usage can be found in the supplementary material.

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

## APPENDIX A: SCATTERING EQUATIONS

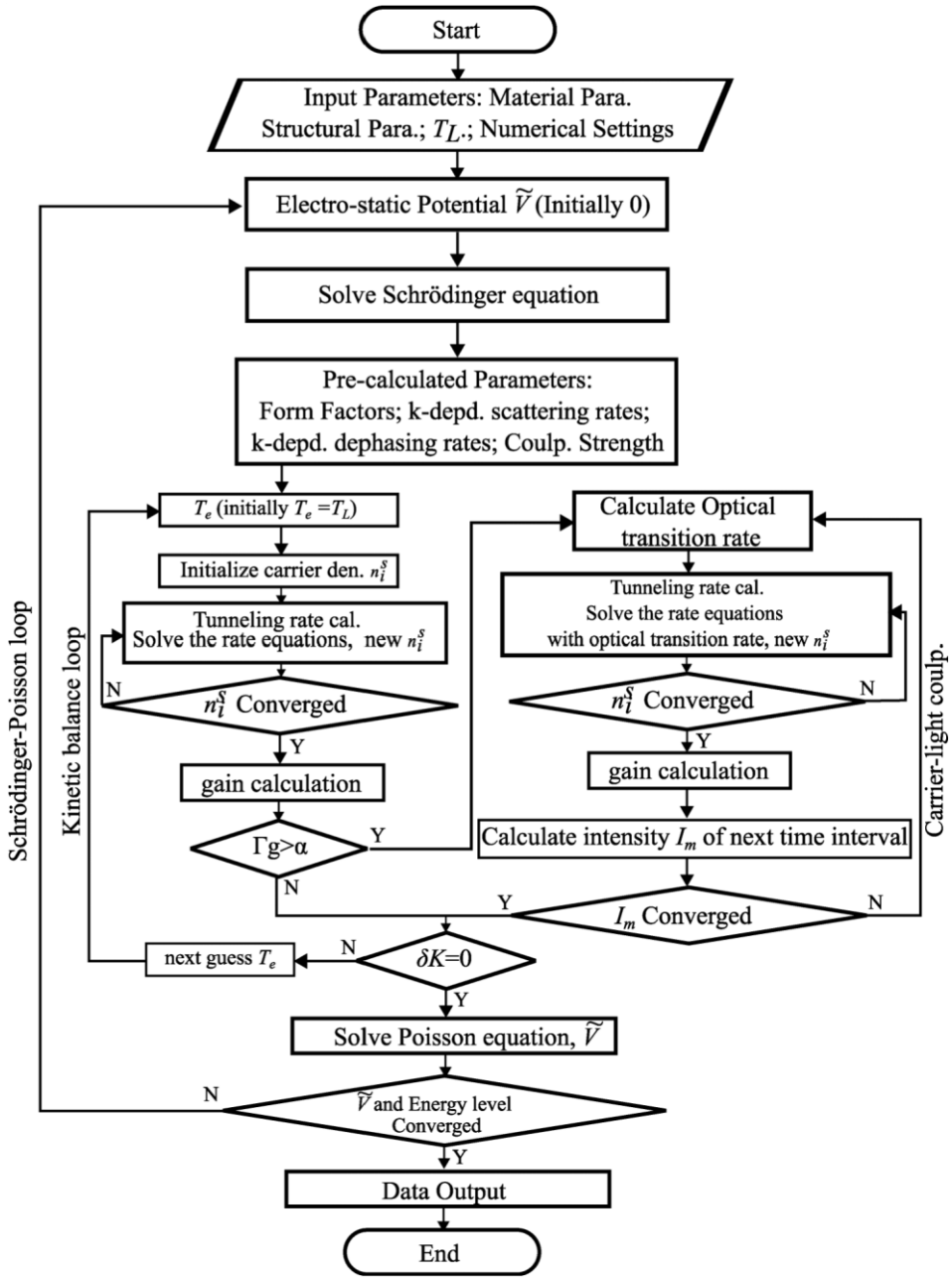



FIG. 12. Simplified flow chart of the model in this work.

The source of the scattering rate equations used in this model and the corresponding equation numbers in the reference are given in Table I.

## APPENDIX B: PROGRAM LAYOUT

The flow chart of the model in this work is given in Fig. 12.

## REFERENCES

[1] R. Köhler, A. Tredicucci, F. Beltram, H. E. Beere, E. H. Linfield, A. G. Davies, D. A. Ritchie, R. C. Iotti, and F. Rossi, "Terahertz semiconductor-heterostructure laser," Nature 417(6885), 156–159 (2002).

[2] F. Wang, X. Qi, Z. Chen, M. Razeghi, and S. Dhillon, "Ultrafast pulse generation from quantum cascade lasers," Micromachines 13(12), 2063 (2022).

[3] F. Wang, X. Qi, V. Pistore, L. Li, G. Agnew, E. Linfield, G. Davies, J. Tignon, J. Mangeney, A. D. Rakić, and S. S. Dhillon, "Ultrafast buildup dynamics of terahertz

pulse generation in mode-locked quantum cascade lasers," Phys. Rev. Appl. 18(6), 064054 (2022).

[4]F. Wang, H. Nong, T. Fobbe, V. Pistore, S. Houver, S. Markmann, N. Jukam, M. Amanti, C. Sirtori, S. Moumdji, R. Colombelli, L. Li, E. Linfield, G. Davies, J. Mangeney, J. Tignon, and S. Dhillon, "Short terahertz pulse generation from a dispersion compensated modelocked semiconductor laser," Laser Photonics Rev. 11(4), 1700013 (2017).

[5]A. Khalatpour, M. C. Tam, S. J. Addamane, J. Reno, Z. Wasilewski, and Q. Hu, "Enhanced operating temperature in terahertz quantum cascade lasers based on direct phonon depopulation," Appl. Phys. Lett. 122(16), 161101 (2023). [6]B. Wen and D. Ban, "High-temperature terahertz quantum cascade lasers," Prog. Quantum Electron. 80, 100363 (2021).

[7]S. Kumar and Q. Hu, "Coherence of resonant-tunneling transport in terahertz quantum-cascade lasers," Phys. Rev. B 80(24), 245316 (2009).

[8]E. Dupont, S. Fathololoumi, and H. C. Liu, "Simplified density-matrix model applied to three-well terahertz quantum cascade lasers," Phys. Rev. B 81(20), 205311 (2010).

[9]R. Terazzi and J. Faist, "A density matrix model of transport and radiation in quantum cascade lasers," New J. Phys. 12(3), 033045 (2010).

[10]B. A. Burnett, A. Pan, C. O. Chui, and B. S. Williams, "Robust density matrix simulation of terahertz quantum cascade lasers," IEEE Trans. Terahertz Sci. Technol. 8(5), 492–501 (2018).

[11]A. Demic, A. Grier, Z. Ikonic, A. Valavanis, C. A. Evans, R. Mohandas, L. Li, E. H. Linfield, A. G. Davies, and D. Indjin, "Infinite-Period density-matrix model for terahertz-frequency quantum cascade lasers," IEEE Trans. Terahertz Sci. Technol. 7(4), 368–377 (2017).

[12]O. Jonasson, "Quantum transport in semiconductor heterostructures using density-matrix and Wigner-function formalisms," Ph.D. thesis (University of Wisconsin-Madison, 2016).

[13]O. Jonasson, F. Karimi, and I. Knezevic, "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). [14]S. Soleimanikahnoj, O. Jonasson, F. Karimi, and I. Knezevic, "Numerically efficient density-matrix technique for modeling electronic transport in mid-infrared quantum cascade lasers," J. Comput. Electron. 20(1), 280–309 (2021).

[15]C. Jirauschek, "Density matrix Monte Carlo modeling of quantum cascade lasers," J. Appl. Phys. 122(13), 133105 (2017).

[16]R. C. Iotti and F. Rossi, "Microscopic theory of quantum-cascade lasers," Semicond. Sci. Technol. 19(4), S323–S326 (2004).

[17]A. Pan, B. A. Burnett, C. O. Chui, and B. S. Williams, "Density matrix modeling of quantum cascade lasers without an artificially localized basis: A generalized scattering approach," Phys. Rev. B 96(8), 085308 (2017).

[18]H. Yasuda, T. Kubis, P. Vogl, N. Sekine, I. Hosako, and K. Hirakawa, "Nonequilibrium Green's function calculation for four-level scheme terahertz quantum cascade lasers," Appl. Phys. Lett. 94(15), 151109 (2009). [19]T. Kubis, C. Yeh, and P. Vogl, "Quantum theory of transport and optical gain in quantum cascade lasers," Phys. Status Solidi (C) 5(1), 232–235 (2008).

[20]T. Kubis, C. Yeh, P. Vogl, A. Benz, G. Fasching, and C. Deutsch, "Theory of nonequilibrium quantum transport and energy dissipation in terahertz quantum cascade lasers," Phys. Rev. B 79(19), 195323 (2009).

[21]P. Greck, S. Birner, B. Huber, and P. Vogl, "Efficient method for the calculation of dissipative quantum transport in quantum cascade lasers," Opt. Express 23(5), 6587 (2015).

[22]A. Kolek and G. Hałdaś, "Optimization of gain region in mid-IR (≈5 μm) QCL," Opt. Express 30(7), 11660 (2022).

[23]H. Yasuda and I. Hosako, "Non-equilibrium Green's function calculation of AlGaAs-well-based and GaSb-based terahertz quantum cascade laser structures," Appl. Phys. Lett. 106(11), 111111 (2015).

[24]T. Grange, D. Stark, G. Scalari, J. Faist, L. Persichetti, L. Di Gaspare, M. De Seta, M. Ortolani, D. J. Paul, G. Capellini, S. Birner, and M. Virgilio, "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). [25]G. Haldas, A. Kolek, and I. Tralle, "Modeling of Mid-infrared quantum cascade laser by means of nonequilibrium Green's functions," IEEE J. Quantum Electron. 47(6), 878–885 (2011).

[26]A. Wacker, M. Lindskog, and D. O. Winge, "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).

[27]M. Franckié, L. Bosco, M. Beck, C. Bonzon, E. Mavrona, G. Scalari, A. Wacker, and J. Faist, "Two-well quantum cascade laser optimization by non-equilibrium Green's function modelling," Appl. Phys. Lett. 112(2), 021104 (2018).

[28]T.-T. Lin, L. Wang, K. Wang, T. Grange, S. Birner, and H. Hirayama, "Over one Watt output power terahertz quantum cascade lasers by using high doping concentration and variable barrier-well height," Phys. Status Solidi RRL 16(7), 2200033 (2022).

[29]T.-T. Lin, L. Wang, K. Wang, T. Grange, and H. Hirayama, "Optimization of terahertz quantum cascade lasers by suppressing carrier leakage channel via high-energy state," Appl. Phys. Express 11(11), 112702 (2018).

[30]K. Wang, T. Grange, T.-T. Lin, L. Wang, Z. Jéhn, S. Birner, J. Yun, W. Terashima, and H. Hirayama, "Broadening mechanisms and self-consistent gain calculations for GaN quantum cascade laser structures," Appl. Phys. Lett. 113(6), 061109 (2018).

[31]P. Harrison, D. Indjin, I. Savić, Z. Ikonić, C. A. Evans, N. Vukmirović, R. W. Kelsall, J. McTavish, V. D Jovanović, and V. Milanović, "On the coherence/incoherence of electron transport in semiconductor heterostructure optoelectronic devices," Proc. SPIE 6909, 690912 (2008).

[32]P. Harrison, and A. Valavanis, Quantum Wells, Wires and Dots: Theoretical and Computational Physics of Semiconductor Nanostructures, 4th ed. (Wiley, Hoboken, NJ, 2016).

[33]K. Donovan, P. Harrison, and R. W. Kelsall, "Self-consistent solutions to the intersubband rate equations in quantum cascade lasers: Analysis of a GaAs/$Al_xGa_{1-x}As$ device," J. Appl. Phys. 89(6), 3084–3090 (2001).

[34]D. Indjin, P. Harrison, R. W. Kelsall, and Z. Ikonić, "Self-consistent scattering theory of transport and output characteristics of quantum cascade lasers," J. Appl. Phys. 91(11), 9019–9026 (2002).

[35]T. Kubis and P. Vogl, "Assessment of approximations in nonequilibrium Green's function theory," Phys. Rev. B 83(19), 195304 (2011).

[36]T. Schmielau and M. F. Pereira, "Impact of momentum dependent matrix elements on scattering effects in quantum cascade lasers: Momentum dependent scattering in QCLs," Phys. Status Solidi (B) 246(2), 329–331 (2009).

[37]A. Hamadou, S. Lamari, and J.-L. Thobel, "Dynamic modeling of a midinfrared quantum cascade laser," J. Appl. Phys. 105(9), 093116 (2009).

[38]Y. Petitjean, F. Destic, J.-C. Mollier, and C. Sirtori, "Dynamic modeling of terahertz quantum cascade lasers," IEEE J. Sel. Top. Quantum Electron. 17(1), 22–29 (2011).

[39]G. Agnew, A. Grier, T. Taimre, Y. L. Lim, M. Nikolić, A. Valavanis, J. Cooper, P. Dean, S. P. Khanna, M. Lachab, E. H. Linfield, A. G. Davies, P. Harrison, Z. Ikonić, D. Indjin, and A. D. Rakić, "Efficient prediction of terahertz quantum cascade laser dynamics from steady-state simulations," Appl. Phys. Lett. 106(16), 161105 (2015).

[40]N. W. Almond, X. Qi, R. Degl'Innocenti, S. J. Kindness, W. Michailow, B. Wei, P. Braeuninger-Weimer, S. Hofmann, P. Dean, D. Indjin, E. H. Linfield, A. G. Davies, A. D. Rakić, H. E. Beere, and D. A. Ritchie, "External cavity terahertz quantum cascade laser with a metamaterial/graphene optoelectronic mirror," Appl. Phys. Lett. 117(4), 041105 (2020).

[41]L. Lever, A. Valavanis, Z. Ikonić, and R. W. Kelsall, "Simulated [111] Si–SiGe terahertz quantum cascade laser," Appl. Phys. Lett. 92(2), 021124 (2008).

[42]A. Valavanis, L. Lever, C. A. Evans, Z. Ikonić, and R. W. Kelsall, "Theory and design of quantum cascade lasers in (111) n-type Si/SiGe," Phys. Rev. B 78(3), 035420 (2008).

[43]J. T. Lü and J. C. Cao, "Monte Carlo simulation of hot phonon effects in resonant-phonon-assisted terahertz quantum-cascade lasers," Appl. Phys. Lett. 88(6), 061119 (2006).

[44]H. Callebaut, S. Kumar, B. S. Williams, Q. Hu, and J. L. Reno, "Analysis of transport properties of tetrahertz quantum cascade lasers," Appl. Phys. Lett. 83(2), 207–209 (2003).

[45]C. Jirauschek and P. Lugli, "Monte-Carlo-based spectral gain analysis for terahertz quantum cascade lasers," J. Appl. Phys. 105(12), 123102 (2009). [46]H. Callebaut and Q. Hu, "Importance of coherence for electron transport in terahertz quantum cascade lasers," J. Appl. Phys. 98(10), 104505 (2005).

[47]R. Terazzi, "Transport in quantum cascade lasers," Ph.D. thesis (ETH Zurich, 2012).

[48]T. Ando, "Line width of inter-subband absorption in inversion layers: Scattering from charged ions," J. Phys. Soc. Jpn. 54(7), 2671–2675 (1985).

[49]T. Unuma, M. Yoshita, T. Noda, H. Sakaki, and H. Akiyama, "Intersubband absorption linewidth in GaAs quantum wells due to scattering by interface roughness, phonons, alloy disorder, and impurities," J. Appl. Phys. 93(3), 1586–1597 (2003).

[50]S. G. Razavipour, "Design, analysis, and characterization of indirectly-pumped terahertz quantum cascade lasers," Ph.D. thesis (University of Waterloo, 2013). [51]T. Miyoshi and D. Ban, "Investigation of Coulomb scattering in terahertz quantum cascade lasers," J. Appl. Phys. 129(15), 153102 (2021).

[52]T. Miyoshi and D. Ban, "Barrier height study of two-well resonant-phonon terahertz quantum cascade lasers. I.: The third-order tunneling current theory," J. Appl. Phys. 130(18), 183103 (2021).

[53]C. Jirauschek and T. Kubis, "Modeling techniques for quantum cascade lasers," Appl. Phys. Rev. 1(1), 011307 (2014).

[54]C. Jirauschek, "Monte Carlo study of carrier-light coupling in terahertz quantum cascade lasers," Appl. Phys. Lett. 96(1), 011103 (2010). [55]R. C. Iotti and F. Rossi, "Nature of charge transport in quantum-cascade lasers," Phys. Rev. Lett. 87(14), 146603 (2001).

[56]J. D. Cooper, "Simulation of surface acoustic wave modulation of quantum cascade lasers," Ph.D. thesis (University of Leeds, 2013).

[57]J. D. Cooper, A. Valavanis, Z. Ikonić, P. Harrison, and J. E. Cunningham, "Finite difference method for solving the Schrödinger equation with band nonparabolicity in mid-infrared quantum cascade lasers," J. Appl. Phys. 108(11), 113109 (2010).

[58]A. Demić, J. Radovanović, and V. Milanović, "Analysis of the influence of external magnetic field on transition matrix elements in quantum well and quantum cascade laser structures," Superlattices Microstruct. 96, 134–149 (2016).

[59]X. Gao, D. Botez, and I. Knezevic, "X-valley leakage in GaAs-based midinfrared quantum cascade lasers: A Monte Carlo study," J. Appl. Phys. 101(6), 063101 (2007).

[60]I. Tan, G. L. Snider, L. D. Chang, and E. L. Hu, "A self-consistent solution of Schrödinger–Poisson equations using a nonuniform mesh," J. Appl. Phys. 68(8), 4071–4076 (1990).

[61]E. Cassan, "On the reduction of direct tunneling leakage through ultrathin gate oxides by a one-dimensional Schrödinger–Poisson solver," J. Appl. Phys. 87(11), 7931–7939 (2000).

[62]S. Fathololoumi, E. Dupont, C. W. I. Chan, Z. R. Wasilewski, S. R. Laframboise, D. Ban, A. Mátyás, C. Jirauschek, Q. Hu, and H. C. Liu, "Terahertz quantum cascade lasers operating up to ~200 K with optimized oscillator strength and improved injection tunneling," Opt. Express 20(4), 3866 (2012).

[63]C. Sirtori, P. Kruck, S. Barbieri, P. Collot, J. Nagle, M. Beck, J. Faist, and U. Oesterle, "GaAs/$Al_xGa_{1-x}As$ quantum cascade lasers," Appl. Phys. Lett. 73(24), 3486–3488 (1998).

[64]A. Mátyás, P. Lugli, and C. Jirauschek, "Photon-induced carrier transport in high efficiency midinfrared quantum cascade lasers," J. Appl. Phys. 110(1), 013108 (2011).

[65]L. Wang, M. Chen, T.-T. Lin, K. Wang, and H. Hirayama, "Interdiffusion limiting on self-consistent optical gain in terahertz quantum cascade lasers," Appl. Phys. Express 16(7), 072004 (2023).

[66]I. Ezhov and C. Jirauschek, "Influence of screening on longitudinal-optical phonon scattering in quantum cascade lasers," J. Appl. Phys. 119(3), 033102 (2016). [67]R. Nelander and A. Wacker, "Temperature dependence and screening models in quantum cascade structures," J. Appl. Phys. 106(6), 063115 (2009).

[68]S.-C. Lee and I. Galbraith, "Intersubband and intrasubband electronic scattering rates in semiconductor quantum wells," Phys. Rev. B 59(24), 15796–15805 (1999).

[69]T. Unuma, T. Takahashi, T. Noda, M. Yoshita, H. Sakaki, M. Baba, and H. Akiyama, "Effects of interface roughness and phonon scattering on intersubband absorption linewidth in a GaAs quantum well," Appl. Phys. Lett. 78(22), 3448–3450 (2001).

[70]E. Dupont, S. Fathololoumi, Z. R. Wasilewski, G. Aers, S. R. Laframboise, M. Lindskog, S. G. Razavipour, A. Wacker, D. Ban, and H. C. Liu, "A phonon scattering assisted injection and extraction based terahertz quantum cascade laser," J. Appl. Phys. 111(7), 073111 (2012).

[71]R. Terazzi, T. Gresch, A. Wittmann, and J. Faist, "Sequential resonant tunneling in quantum cascade lasers," Phys. Rev. B 78(15), 155328 (2008). [72]W. Freeman, "Self-consistent calculation of dephasing in quantum cascade structures within a density matrix method," Phys. Rev. B 93(20), 205301 (2016). [73]V. Rindert, E. Önder, and A. Wacker, "Analysis of high-performing terahertz quantum cascade lasers," Phys. Rev. Appl. 18(4), L041001 (2022).

[74]L. Bosco, M. Franckié, G. Scalari, M. Beck, A. Wacker, and J. Faist, "Thermoelectrically cooled THz quantum cascade laser operating up to 210 K," Appl. Phys. Lett. 115(1), 010601 (2019).

[75]A. Khalatpour, A. K. Paulsen, C. Deimert, Z. R. Wasilewski, and Q. Hu, "High-power portable terahertz laser systems," Nat. Photonics 15(1), 16–20 (2021). [76]V. Spagnolo, G. Scamarcio, H. Page, and C. Sirtori, "Simultaneous measurement of

the electronic and lattice temperatures in GaAs/Al0.45Ga0.55As quantum-cascade lasers: Influence on the optical performance," Appl. Phys. Lett. 84(18), 3690–3692 (2004).

[77]P. Harrison, D. Indjin, and R. W. Kelsall, "Electron temperature and mechanisms of hot carrier generation in quantum cascade lasers," J. Appl. Phys. 92(11), 6921–6923 (2002).

[78]M. S. Vitiello, G. Scamarcio, V. Spagnolo, B. S. Williams, S. Kumar, Q. Hu, and J. L. Reno, "Measurement of subband electronic temperatures and population inversion in THz quantum-cascade lasers," Appl. Phys. Lett. 86(11), 111115 (2005). [79]P. Slingerland, C. Baird, and R. H. Giles, "Application of multi-subband selfconsistent energy balance method to terahertz quantum cascade lasers," Semicond. Sci. Technol. 27(6), 065009 (2012).

[80]Z. Ikonić, P. Harrison, and R. W. Kelsall, "Self-consistent energy balance simulations of hole dynamics in SiGe⁄SiTHz quantum cascade structures," J. Appl. Phys. 96(11), 6803–6811 (2004).

[81]C. Jirauschek, A. Matyas, and P. Lugli, "Modeling bound-to-continuum terahertz quantum cascade lasers: The role of Coulomb interactions," J. Appl. Phys. 107(1), 013104 (2010).

[82]A. Demic, "Density matrix modelling of terahertz frequency quantum cascade lasers: Steady state analysis and Maxwell-Bloch dynamics," Ph.D. thesis (University of Leeds, 2019).

[83]O. Bonno, J.-L. Thobel, and F. Dessenne, "Modeling of electron–electron scattering in Monte Carlo simulation of quantum cascade lasers," J. Appl. Phys. 97(4), 043702 (2005). [84]

G. Scalari, M. I. Amanti, C. Walther, R. Terazzi, M. Beck, and J. Faist, "Broadband THz lasing from a photon-phonon quantum cascade structure," Opt. Express 18(8), 8043 (2010).

[85]S. Fathololoumi, E. Dupont, Z. R. Wasilewski, C. W. I. Chan, S. G. Razavipour, S. R. Laframboise, S. Huang, Q. Hu, D. Ban, and H. C. Liu, "Effect of oscillator strength and intermediate resonance on the performance of resonant phononbased terahertz quantum cascade lasers," J. Appl. Phys. 113(11), 113109 (2013).

[86]R. S. Dhar, S. G. Razavipour, E. Dupont, C. Xu, S. Laframboise, Z. Wasilewski, Q. Hu, and D. Ban, "Direct nanoscale imaging of evolving electric field domains in quantum structures," Sci. Rep. 4(1), 7183 (2014).

[87]I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band parameters for III–V compound semiconductors and their alloys," J. Appl. Phys. 89(11), 5815–5875 (2001).

[88]M. A. Belkin, J. A. Fan, S. Hormoz, F. Capasso, S. P. Khanna, M. Lachab, A. G. Davies, and E. H. Linfield, "Terahertz quantum cascade lasers with copper metal-metal waveguides operating up to 178 K," Opt. Express 16(5), 3242 (2008).

[89]D. O. Winge, E. Dupont, and A. Wacker, "Ignition of quantum cascade lasers in a state of oscillating electric field domains," Phys. Rev. A 98(2), 023834 (2018).

[90]H. Sakaki, T. Noda, K. Hirakawa, M. Tanaka, and T. Matsusue, "Interface roughness scattering in GaAs/AlAs quantum wells," Appl. Phys. Lett. 51(23), 1934–1936 (1987).

[91]H. Willenberg, G. H. Döhler, and J. Faist, "Intersubband gain in a Bloch oscillator and quantum cascade laser," Phys. Rev. B 67(8), 085315 (2003).