Importance of coherence for electron transport in terahertz quantum cascade lasers

Hans Callebaut and Qing Hu

Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

A density matrix approach is used in combination with a tight-binding model to describe electron transport in terahertz quantum cascade lasers and is incorporated into a Monte Carlo simulation. Scattering events, including LO-phonon, electron-electron, and ionized impurity scattering, are treated semiclassically but contribute to dephasing scattering. In addition, a phenomenological "pure dephasing rate" was introduced to take into account dephasing caused by interface roughness scattering. This model was used to investigate the influence of dephasing on electron transport through a barrier. Additionally, current densities, populations and electron temperatures were calculated for a simple three-level structure and a five-level structure that achieved lasing at 3.2 THz, and the results were compared to a semiclassical simulation. We find that the inclusion of coherent transport and dephasing in the calculations is essential when transport is dominated by transitions between weakly coupled states. © *2005 American Institute of Physics.*
[DOI: 10.1063/1.2136420]

I. INTRODUCTION

In quantum cascade lasers (QCLs), resonant tunneling is a critical transport mechanism, and as such it is the subject of active theoretical and experimental research.[1–3] However, although a qualitative understanding is straightforward, it is not always clear how to quantify the exact effect of coherent and incoherent transport. The effects of resonant tunneling and dephasing are most important when describing the transport between two weakly coupled energy states, i.e., tunneling through a thick barrier such as an injector barrier. In the calculation and analysis of QCLs, the localization of wavefunctions due to dephasing scattering is often disregarded, which can lead to unphysical results and limit the usefulness of the simulation.[4] Therefore, it is necessary to include a model for sequential tunneling to analyze the electron transport in QCLs over a broad bias range. Although simulations using nonequilibrium Green (NEG) function analysis[5] have shown some promising results recently, the complexity and computational burden of this method limit its utility in obtaining an intuitive picture of electron transport. On the other hand, the density matrix formalism provides an easily accessible description of coherent electron transport, and is widely used to model optical and electronic transitions.[6] To investigate the importance of coherent transport in THz QCLs, we used Monte Carlo (MC) simulations with a semiclassical approach and with a density matrix (MC-DM) approach to model two experimentally tested devices, a three-level superlattice[7] and a 3.2 THz QCL (Ref. 8) that has five levels in each module participating in the transport.

The importance of coherent transport in multiple quantum well (MQW) structures can be appreciated by considering a simple superlattice as shown in Fig. 1. In this simple structure, only two energy levels in each well, 1 and 2, participate in the transport process as electrons move from the left to the right under an applied electric field. Figure 1(a) illustrates the scheme that is described by the semiclassical model, in which the entire superlattice is treated as a single quantum mechanical system with a well-defined Hamiltonian. All the subband energy levels are eigenstates (which are stationary by definition) of this Hamiltonian. The transport process is the collective effect of intersubband scattering between the various subbands (eigenstates) involved, and can be calculated using the Fermi's golden rule approximation. This is essentially a rate-equation approach and there is no coherent oscillatory time evolution among the subband electron populations. In this picture, the electron wavefunctions always correspond to the stationary eigenstates, and scattering transports an electron from one eigenstate to another. Under resonant bias conditions, the ground state $1'$ in one well is aligned with the excited level 2 of the adjacent well. These two levels form a spatially extended doublet with the "symmetric" wavefunction $|S\rangle$ as the lower-energy state, and the "antisymmetric" wavefunction $|A\rangle$ as the higher-energy state [Fig. 1(a)]. The energy separation of the doublet is the anticrossing gap $\Delta_{1'2}$. In this semiclassical picture, the transport through an energy barrier is effectively instantaneous, as both $|S\rangle$ and $|A\rangle$ are spatially extended across the barrier, and consequently the barrier causes no "resistance" to the electron transport under the resonant bias. The only "bottleneck" of this transport process is the energy-relaxing (inelastic) intersubband scattering from the doublet $|S\rangle$ and $|A\rangle$ into 1, or equivalently (in the case of a superlattice) into the doublet $|S''\rangle$ and $|A''\rangle$ formed by 1 and $2''$ of the following well (not shown in the figure). As a result, the current density under this resonant bias is independent of the barrier thickness, which is only valid in the absence of dephasing, and is thereby unphysical for real devices.

In contrast, in Fig. 1(b) localized basis states are used and electron transport through the barrier takes place via a coherent time evolution of these states, i.e., it takes the elec-
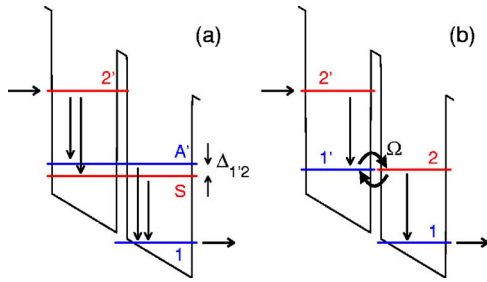
FIG. 1. Difference between semiclassical and coherent picture of coupled quantum wells. (a) Semiclassical picture. The wavefunctions represent eigenstates of the Hamiltonian and are delocalized at resonance. Transport through the barrier happens as soon as electrons enter levels $|S\rangle$ or $|A\rangle$. (b) Coherent picture. The wave packet is initially localized in the left well. Electrons are transported through the barrier with Rabi oscillations at frequency Ω due to the interaction between $1'$ and 2.

trons a finite time to get from one well to the next. In this scheme, at an initial time, the electron wave packet resides at the bottom of the left well in state $1'$. This wave packet can be composed as a coherent superposition of level $|S\rangle$ and level $|A\rangle$. As time evolves, this wave packet oscillates across the middle barrier at the Rabi oscillation frequency $\Delta_{1'2}/\hbar$. In the absence of pure dephasing, this oscillation will be damped only by the intersubband scattering as the wave packet is depleted each time it is in the right well where intersubband scattering takes place. The main bottleneck of the electron transport is again the intersubband scattering lifetime, as in the scenario of Fig. 1(a), even though a finite transport time (half of the Rabi oscillation period) across the barrier increases the dwell time in each well and consequently reduces the current density somewhat. Note that, in the absence of dephasing, the wave packet spends half of its time in either well, so that the time-average of the population distribution is in agreement with the picture described in Fig. 1(a). The most significant difference between the two schemes, however, becomes clear in the presence of dephasing scattering that may be caused by various elastic intrasubband scattering mechanisms, such as interface roughness and electron-impurity scattering. With dephasing scattering, the Rabi oscillation can be damped even in the absence of inelastic intersubband scattering. One may envision that the dephasing scattering can be so strong that the Rabi oscillation is overdamped, i.e., the time evolution of the wave packet from the left to the right well is no longer oscillatory (which is a direct analogy to an overdamped harmonic oscillator). In this strong dephasing limit, the bottleneck of the current transport is the tunneling barrier, which is the scenario discussed by Luryi.[9] Now the time-averaged population distribution will be different in both wells, as electrons pile up behind the barrier, and the simple semiclassical picture of Fig. 1(a) is no longer a good approximation.

So far, most of the analysis of transport processes in QCLs has been based on the semiclassical model described in Fig. 1(a).[4,10–12] This is mainly because QCLs were first developed at midinfrared frequencies, where the photon energy $\hbar\omega > 100$ meV. Consequently, the injection barriers are relatively thin, which results in a large anticrossing gap of $\Delta_{1'2} \sim 10$ meV. Dephasing due to intrasubband scattering does not cause a significant damping to the fast $1' \leftrightarrow 2$ oscil-

lation, and the main bottleneck of the transport is due to intersubband scattering. In fact, this Rabi oscillation at 2.5 THz (~ 10 meV) across the injection barrier has been experimentally observed in a midinfrared QCL by using a time-resolved pump-and-probe method.[2] In THz QCLs, however, the photon energy is much smaller, $\hbar\omega \sim 10-20$ meV. Therefore, the injection barrier must be made thicker with a smaller anticrossing gap of $\Delta_{1'2} \sim 1$ meV in order to maintain a high injection selectivity. In comparison, the dephasing rate, which can be estimated from the measured spontaneous emission linewidth, is relatively higher ($\sim 4-6$ meV in our THz QCLs based on resonant LO-phonon scattering).[13] As a result of this much stronger dephasing relative to the injection anticrossing gap, we have found that transport analysis based on the semiclassical model is quite inadequate. For example, the measured maximum current densities at resonance are observed to be very sensitive to the thickness of the injection barriers. Also, simulations based on the semiclassical model tend to overestimate the current densities and material gains in our laser devices,[4,10] and even predict substantial levels of gain in experimental devices that did not achieve lasing. It is this significant discrepancy between simulation results based on the semiclassical model and experimental results that motivates us to pursue the investigation described in this paper: the importance of the coherent aspect of the transport process, or equivalently, the quantitative effect of dephasing scattering on the transport process involving subband levels at resonance.

In the following sections, we will briefly introduce the density matrix formalism and its implementation in the MC simulation. We will then compare the semiclassical and density matrix approaches by investigating electron transport through a barrier in resonant and nonresonant bias conditions, and use the density matrix MC simulation to investigate a simple three-level structure and a five-level QCL that achieved lasing at 3.2 THz.

II. THEORETICAL APPROACH

To describe the time evolution and phase coherence of a large number of particles, we can choose from several different approaches. The most straightforward method would be to use a Schrödinger picture, where we keep track of the full wavefunction of every particle. From these wavefunctions we can then easily find the relevant macroscopic quantities, like current, population density or optical gain by summing over the contributions from each particle. However, due to the inherently statistical nature of these quantities, much of the information contained in those wavefunctions is averaged out and turns out not to be relevant for the macroscopic picture.

Another, more efficient approach is the density matrix formalism,[14] which describes the statistical distribution of quantum states in a system. This method allows us to treat the properties of a large ensemble of electrons (particles) statistically, without worrying about the exact details of the individual electrons' wavefunctions. A generic particle from this ensemble can be represented by a wavefunction ψ:

$$|\psi\rangle = \sum_i c_i |\phi_i\rangle, \tag{1}$$

where ϕ_i are the basis wavefunctions belonging to the Hamiltonian H_0 of the unperturbed system, and

$$c_i = \langle \phi_i | \psi \rangle. \tag{2}$$

We can then define the density operator

$$\rho(t) = |\psi(t)\rangle\langle\psi(t)|, \tag{3}$$

which takes the form of a projection operator. The density operator can be interpreted as a description of the probability distribution in a system. For an ensemble of particles, the density matrix elements are defined as the ensemble averages:

$$\rho_{mn} = \langle c_n c_m^* \rangle. \tag{4}$$

The diagonal elements ρ_{ii} describe the probability of finding the system in state $|i\rangle$ and are proportional to the population density of that state. The off-diagonal elements ρ_{ij} are related to the polarization between states i and j and describe the degree of coherent interaction. Consistent with its interpretation as "probability matrix," it can be shown that

$$\mathrm{Tr}(\rho) = \sum_i^N \rho_{ii} = 1, \tag{5}$$

which reflects that the total population density is conserved. Another important property is that the magnitude of each off-diagonal element is smaller than or equal to the geometric mean of the corresponding diagonal elements (Schwartz inequality):

$$\rho_{ii}\rho_{ij} \geqslant |\rho_{ij}|^2. \tag{6}$$

Physically, the equality corresponds to a "pure state" described by a single wavefunction, such as the one described by Eqs. (1)–(3). Note that this pure state needs not represent just a single electron, but can instead also be used to describe time evolution of an ensemble of particles. On the other hand, the inequality in Eq. (6) refers to a "mixed state," which can in general be broken up into simpler constituent pure states:

$$\rho^{\mathrm{mixed}} = \sum_i |\psi_i\rangle\langle\psi_i|. \tag{7}$$

A mixed state cannot be described with a single wavefunction, and represents an ensemble consisting of independently evolving pure states, i.e., a mixed state is in fact an ensemble of ensembles. In essence, a mixed state reflects the interaction between a subsystem that is characterized by a well-defined Hamiltonian H_0 and the rest of the environment, whose effects are too complex to be dealt with from first principles. The effect of dephasing can then be considered as the scrambling of the phase coherence of some electrons in one of the constituent pure state ensembles. In this picture, dephasing causes the involved electrons to be removed from their original pure state, and subsequently added back to the mixed state in a new constituent pure state, but with a phase unrelated to its original phase. The net effect is that the population remains unaffected (diagonal elements ρ_{ii}) whereas the average coherence ρ_{ij} decreases due to the randomization. As will be explained in more detail later, the density matrices used in the MC simulation will generally be mixed states rather than pure states, since they describe the coherence of all electrons with the same transverse momentum $\mathbf{k}_\perp$.

To capture the dynamics of coherent transport we need the equation of motion, which describes the time-evolution of the density operator and hence of the populations and polarizations. This equation of motion is also known as the quantum Liouville equation and can be written as

$$\frac{\partial\rho}{\partial t} = -\frac{i}{\hbar}[H,\rho], \tag{8}$$

with $H = H_0 + H'$, and H' represents a perturbation. In our case, H' consists only of an adjustment of the electron potential ΔV_{TB} due to the coupling of the localized states in one QCL module to the states in the neighboring modules, in the spirit of the tight-binding (TB) model. It is possible to rewrite the above equation so that its expression is formally identical to the calculation of a wavefunction. Looking at the right-hand side of Eq. (8), we see that the elements of the density operator ρ_{ij} undergo a linear transformation and it is therefore possible to define a linear operator $\mathcal{L}$ to describe this transformation. The Liouville Eq. (8) then becomes

$$\frac{\partial\rho}{\partial t} = -\frac{i}{\hbar}\mathcal{L}\rho, \tag{9}$$

where $\mathcal{L}$ is called the Liouville operator, and

$$\mathcal{L}_{ij,mn} = H_{im}\delta_{jn} - H_{jn}^*\delta_{im}. \tag{10}$$

In this representation (Liouville space), the superoperator $\mathcal{L}$ is a $N^2 \times N^2$ matrix (N is the number of states in the system) and ρ is a N^2-dimensional vector. The number of elements in $\mathcal{L}$ scales with the fourth power of N, and systems with many states can quickly pose almost insuperable computational challenges. The simplifications and approximations discussed in Sec. III effectively "remove" many off-diagonal elements, and vastly simplify the numerical implementation. Note that Eq. (9) looks very similar to the Schrödinger equation, and this allows us to apply the same formalisms to both Hilbert (wavefunction) and Liouville (density operator) spaces. Using this complete formal analogy, we can apply the calculation techniques developed for wavefunctions to the density matrix formalism and obtain the desired results.

So far we have described the coherent time evolution of an electron wavefunction in a system H_0 with only a constant perturbation ΔV_{TB} due to the interaction with the neighboring modules. Note that in the absence of scattering implicitly assumed in Eq. (9), the transverse momentum $\mathbf{k}_\perp$ is conserved and ΔV_{TB} is nonzero only for states with an identical $\mathbf{k}_\perp$. There is no coherent interaction between states with different in-plane momentum in this approximation.[15] As explained in the following, transport between states with different $\mathbf{k}_\perp$ is handled separately, through semiclassical scattering mechanisms within the same module. A fully coherent description of the many interactions, such as electron-phonon, electron-impurity and electron-electron scattering,

would be very involved and computationally intensive. Therefore, it is more convenient to describe them as semiclassical scattering events, in which electron scattering rates are described by Fermi's golden rule. In view of the setup of our model, this is a reasonable approximation. In THz QCLs, the injection barrier is usually much thicker than the other (intramodule) barriers, which results in the "intermodule" anticrossing gaps (where each level is on a different side of the injection barrier) being considerably smaller than "intramodule" anticrossing gaps. As explained in greater detail in Sec. IV, the intermodule interactions are thus more sensitive to dephasing, and are best described with Eq. (9), whereas for the intramodule transitions a semiclassical model is adequate. The scattering events transport electrons between different states within the system, and hence cause the relaxation of the population with a scattering time T_1. In addition to this relaxation scattering, we can also consider "pure dephasing" events that merely scramble the phase correlation between two states at a rate T_2^{-1} without causing depopulation. As mentioned before, pure dephasing accounts for the effects of the bath on the electrons in the system, and as such describe scattering events that are not explicitly included in the simulation model. Both relaxation scattering and pure dephasing contribute to the dephasing time τ_{deph}:

$$\frac{1}{\tau_{\text{deph}}} = \frac{1}{2T_1} + \frac{1}{T_2}. \qquad (11)$$

We note that the contribution of the relaxation scattering is half of that of the pure dephasing. This is due to the fact that T_2 describes the relaxation of the polarization ρ_{ij}, which is proportional to the amplitude of the oscillation ($\rho_{ij} \propto e^{-t/T_2}$); on the other hand, T_1 is a probability decay rate that reduces $\rho_{ii} \propto |\rho_{ij}|^2$, proportional to the energy density ($\rho_{ii} \propto e^{-t/T_1} \rightarrow \rho_{ij} \propto e^{-t/2T_1}$). In the equations of motion, scattering and pure dephasing add extra relaxation terms to the expressions for ρ_{ii} and ρ_{ij}, which can be incorporated in the density matrix formalism with a corresponding superoperator $\mathcal{F}$. Eq. (9) now becomes

$$\frac{\partial \rho_{ij}}{\partial t} = \sum_{m,n} (\mathcal{L}_{ij,mn} + \mathcal{F}_{ij,mn}) \rho_{mn}, \qquad (12)$$

with

$$\mathcal{F}_{ij,mn} = -\frac{1}{2}(\gamma_i + \gamma_j)\delta_{im}\delta_{jn} - \Gamma_{ij}^{\text{pure}}\delta_{im}\delta_{jn}, \quad i \neq j, \qquad (13)$$

and

$$\mathcal{F}_{ii,jj} = \gamma_{ji}(1 - \delta_{ij}) - \gamma_i \delta_{ij}. \qquad (14)$$

Here γ_i and γ_j correspond to the total scattering rates out of the i and j levels, γ_{ji} is the net scattering rate from level j to level i and $\Gamma_{ij}^{\text{pure}}(=T_2^{-1})$ is the pure dephasing rate of the coherent transport between i and j. It is important to note that the dephasing of ρ_{ij} is not due only to scattering between i and j, but rather to *all* scattering events involving either i or j. Electrons scattering out of a level i disrupt the coherent transport from and to i, and as such dephase all polarizations $\rho_{ik}(k \neq i)$ that involve level i. In Eq. (12), all the stochastic aspects of the electron transport are included in the operator $\mathcal{F}$. Using a phenomenological pure dephasing time constant T_2, Eq. (12) can be solved analytically to yield an expression of current density.[16,17] In this work, we use a MC method to deal with the effect of $\mathcal{F}$ numerically, thus allowing us to investigate more complicated structures than those in Refs. 16 and 17. Details of the numerical implementation of the density matrix MC (DM-MC) simulation are given in the following section.

III. NUMERICAL IMPLEMENTATION

The MC method is very flexible and allows for a relatively simple and straightforward simulation of the equations of motion (12). On the other hand, the introduction of the density matrix formalism requires that we keep track of many more variables (e.g., the polarizations) than in the semiclassical case, and the computational requirements of a full-fledged implementation of these equations rapidly become very demanding. However, the proper choice of the basis wavefunctions allows us to develop an intuitive description of the transport in a QCL, which is in part coherent, and in part semiclassical.

In most of the demonstrated THz QCL structures, the levels within one module couple more strongly to each other than to levels in a different module, and a semiclassical description of transport within these regions is adequate. In order to reduce complexity, we restrict the use of coherent transport to model the transport between those modules, e.g., through the injector barrier, while retaining a semiclassical description for the transport inside each module. In this picture, current bottlenecks are described with the coherent density matrix model, whereas the transport through the rest of the device is described semiclassically. The basic implementation of this semiclassical part of the simulation was described previously,[4,10,18] and includes semiclassical electronphonon (acoustic and LO), electron-impurity (e-imp) and electron-electron (e-e) scattering. A nonequilibrium, multisubband screening model was used for e-imp and e-e interactions.

As outlined in the introduction, the choice of the proper basis wavefunctions can be very important. This basis is used to calculate semiclassical scattering rates as well as to compose localized wave packets to model tunneling behavior and localization. For reasons that are explored in Sec. IV, the use of spatially extended wavefunctions (as in semiclassical simulations) as basis wavefunctions fails to reproduce the experimentally observed tunneling behavior for transport through a barrier. To provide an intuitive picture of the electron states involved in transport, we chose a basis of wavefunctions which were localized within a module or submodule of the QCL under investigation. The thick injector barriers that confine a module form an obstruction for the electron transport, and we expect dephasing effects to be most prominent there. The choice of the basis wavefunctions as confined to either side of this barrier, makes it easy to describe and calculate resonant tunneling.

To find the localized wavefunctions ϕ_i^0, we consider a single, isolated module under bias, embedded in material

with the same composition as the barriers. The wavefunctions in the previous ($n=-1$) and subsequent ($n=1$) modules are then found from

$$\phi_i^n(x) = \phi_i^0(x - nl_{\text{mod}}) \tag{15}$$

with energy $E_i^n = E_i^0 + nqV_{\text{bias}}$, where $q=-1.6\times10^{-19}C$ is the electron charge, l_{mod} is the module length, and V_{bias} is the applied voltage per module. The interaction ΔV_{ij} between the localized wavefunctions ϕ_i^m and ϕ_j^n with respective in-plane momenta $\mathbf{k}_\perp = \mathbf{k}_i$ and $\mathbf{k}_j$ is then determined with a tight-binding model:

$$\Delta V_{ij}^n(\mathbf{k}_i, \mathbf{k}_j) = \int_{-\infty}^{+\infty} \phi_i^0(z)\Delta V(z)\phi_j^n(z)dz\,\delta_{\mathbf{k}_i\mathbf{k}_j}. \tag{16}$$

Here $\Delta V(z)$ is the difference in the confining potential between a single module and a superlattice composed of a repetition of this module:

$$\Delta V(z) = V_{\text{barr}} - \sum_{m<0} V_{\text{mod}}(z - ml_{\text{mod}}), \quad z < 0,$$

$$\Delta V(z) = 0, \quad 0 < z < l_{\text{mod}}, \tag{17}$$

$$\Delta V(z) = V_{\text{barr}} - \sum_{m>0} V_{\text{mod}}(z - ml_{\text{mod}}), \quad z > l_{\text{mod}},$$

where V_{barr} is the barrier potential and V_{mod} is the unbiased module potential profile. In practice, $\Delta V(z)$ is restricted to the influence of next-neighbor modules ($n=-1,1$) because QCL modules are usually very wide compared to the extinction length of the localized wavefunctions. Also note that $\Delta V_{ij}^1 = \Delta V_{ji}^{-1}$. The coherent intramodule transport due to ΔV_{ij}^0 is negligible compared to the scattering-mediated transport. It should be stressed again that in the absence of scattering, the transverse momentum $\mathbf{k}_\perp$ is conserved and $\Delta V_{ij}^n(\mathbf{k}_i, \mathbf{k}_j)$ is nonzero only for states with an identical $\mathbf{k}_\perp$. In what follows we will abbreviate $\Delta V_{ij}^1(\mathbf{k}_i, \mathbf{k}_j)\delta_{\mathbf{k}_i\mathbf{k}_j}$ with ΔV_{ij}. Note that this calculated interaction depends (weakly) on bias, as the changing potential and interactions within the module shift the wavefunctions around.

It should be pointed out that this approach yields a direct estimate of the anticrossing gap $\Delta_{ij} = 2|\Delta V_{ij}|$ between levels ϕ_i^0 and ϕ_j^1, and it makes this parameter easily accessible for investigation. Since the injector barrier thickness only affects the calculated Δ_{ij} and not the localized wavefunctions, this approach is very convenient to study the effect of barrier thickness on transport (see Sec. IV).

As mentioned before, we choose to adopt a "hybrid" strategy when including the density matrix formalism into the MC simulation. Only the transport through the injector barriers is modeled in the DM formalism, the transport through the rest of the module is still handled semiclassically. This means that all scattering rates are calculated using Fermi's golden rule instead of a full density matrix description, and are simulated using a MC approach. The transport through the barrier is handled by the quantum Liouville equation [Eqs. (12)–(14)] which includes the depopulation and pure dephasing scattering rates in the matrix $\mathcal{F}$. The solution to this equation exhibits oscillations on timescales

varying from a few femtoseconds to tens of picoseconds, and a full MC implementation of these equations would be very computationally intensive. Instead, we choose to analytically calculate the solutions to the equation of motion with a global phenomenological pure dephasing rate $\Gamma_{ij}^{\text{pure}} = T_2^{-1}$ that applies to all subband states:

$$\frac{\partial\rho_{ij}}{\partial t} = \sum_{m,n}[\mathcal{L}_{ij,mn} - T_2^{-1}\delta_{im}\delta_{jn}]\rho_{mn}. \tag{18}$$

T_2 can be estimated from measurements of the spontaneous emission linewidth. By doing this, we describe the damping of the Rabi oscillation with two numerical methods: pure dephasing (T_2) is accounted for analytically and is assumed to be constant for the duration of the simulation; scattering dephasing (due to e-LO-phonon, e-e, and e-imp scattering) is calculated using a MC method and can (and will) vary with time. This scheme allows us to separate the time scales of pure and scattering dephasing, as the MC sampling rate is not affected by the pure dephasing time. The description of the time evolution of (an ensemble of) electrons between scattering events includes both coherent transport and pure dephasing. The practical implementation is straightforward. An arbitrary density matrix ρ can be described at time $t=t_0$:

$$\rho(t_0) = \sum_{k,l} c_{kl}(t_0)\rho_{kl}^0, \tag{19}$$

with

$$\rho_{kl}^0 = |\phi_k\rangle\langle\phi_l|. \tag{20}$$

Because the Liouville operator is a linear operator, we can describe the time dependence of a density matrix as the sum of the time evolutions of its components. If we write $\bar{\rho}_{kl}(t)$ for the solution to the equation of motion (including only pure dephasing scattering as discussed previously), i.e., $\bar{\rho}_{kl}(t)$ is solved from Eq. (18) for the basis density matrices ρ_{kl}^0, we find

$$\bar{\rho}_{kl}(t) = \sum_{m,n} c_{mn}^{kl,0}(t)\rho_{mn}^0, \tag{21}$$

and after a flight time Δt, ρ is transformed into

$$\rho(t_0 + \Delta t) = \sum_{k,l,m,n} c_{kl}(t_0)c_{mn}^{kl,0}(\Delta t)\rho_{mn}^0. \tag{22}$$

The *a priori* unknown dephasing rates due to relaxation scattering can then be added during the MC simulation by setting the appropriate off-diagonal element ρ_{ij} to zero every second time a scattering event affecting levels i or j happens. This ensures that the dephasing rate due to relaxation scattering is half of $(\gamma_i + \gamma_j)$, as seen in Eq. (13). The affected diagonal element ρ_{ii} is adjusted accordingly as in the case of the semiclassical model.

A semiclassical MC simulation deals with "integer" particles, i.e., every simulated particle represents an ensemble of electrons that is not broken up during the course of the simulation. In principle, all the particles in the ensemble can be name tagged and monitored during their transport process through the whole structure. The ensemble evolves and scatters as a whole, and at every point in time has a single well-defined momentum $\mathbf{k}_\perp$. Whether a particle is viewed as a

single electron or an ensemble of electrons is not important, since it does not affect the particle dynamics. Because the particles are indivisible, conservation laws dictate that their total number remains constant over the course of the simulation.

In the density matrix MC simulation, however, this is no longer the case, and we can no longer identify a particle with a single electron. Instead, each particle needs to be treated as an ensemble of electrons. At the beginning of the simulation, there is a limited number of particles with a specific energy and momentum, each represented by a density matrix. The absence of coherent interaction between two states with different $\mathbf{k}_\perp$, makes it convenient to assign a different density matrix to every point in $\mathbf{k}_\perp$ space as their (coherent) time evolutions are independent. An ensemble ρ initially localized in one level i will quickly spread out across multiple levels due to coherent transport, while retaining its localization in $\mathbf{k}_\perp$ space:

$$\rho(0) = |\phi_i\rangle\langle\phi_i| \rightarrow \rho(t) = \sum_{k,l} c_{kl}(t)|\phi_k\rangle\langle\phi_l|. \tag{23}$$

However, relaxation scattering is calculated semiclassically and describes the transition of electrons from level i with initial momentum $\mathbf{k}_i$ (in the ensemble $\rho^{\mathbf{k}_i}$) to level j with final momentum $\mathbf{k}_j$. Consequently, a scattering event only affects the parts of $\rho^{\mathbf{k}_i}$ that refer to level i, namely the population density $\rho_{ii}^{\mathbf{k}_i}$ and the coherences $\rho_{ik}^{\mathbf{k}_i}(k \neq i)$, whereas the other elements remain the same:

$$\rho_{ii}^{\mathbf{k}_i} \rightarrow 0, \tag{24}$$

$$\rho_{ii}^{\mathbf{k}_i}, \rho_{ki}^{\mathbf{k}_i}(k \neq i) \rightarrow 0(50\% \text{ chance}). \tag{25}$$

The scattered electrons generally have a different momentum $\mathbf{k}_j \neq \mathbf{k}_i$, and need to be represented in a new density matrix $\rho^{\mathbf{k}_j}$, which initially consists of an electron wave packet localized in level j:

$$\rho^{\mathbf{k}_j} = |\phi_j\rangle\langle\phi_j|. \tag{26}$$

The electron population originally represented by one density matrix is now spread over two density matrices $\rho^{\mathbf{k}_i}$ and $\rho^{\mathbf{k}_j}$, and subsequent scattering will fragment them even further. The result is an ever increasing number of ensembles with different weights, spread out over $\mathbf{k}_\perp$ space. To counter this unbounded proliferation, we chose to group different ensembles according to their distribution in $\mathbf{k}_\perp$ space, by assigning them to "bins" chosen to represent a grid in $\mathbf{k}_\perp$ space. The different density matrices in one bin are very close in $\mathbf{k}_\perp$, and can be approximately described with a single density matrix which is the weighted sum of all density matrices within this bin.

IV. RESULTS AND DISCUSSION

To quantitatively illustrate the effects of dephasing and coherence on transport through a barrier, it is instructive to investigate a simple superlattice structure as sketched in Fig. 2(a), consisting of a succession of coupled quantum wells, separated by a barrier with thickness t_{barr}, and with two energy levels 1 and 2 in each well. Every well corresponds to a
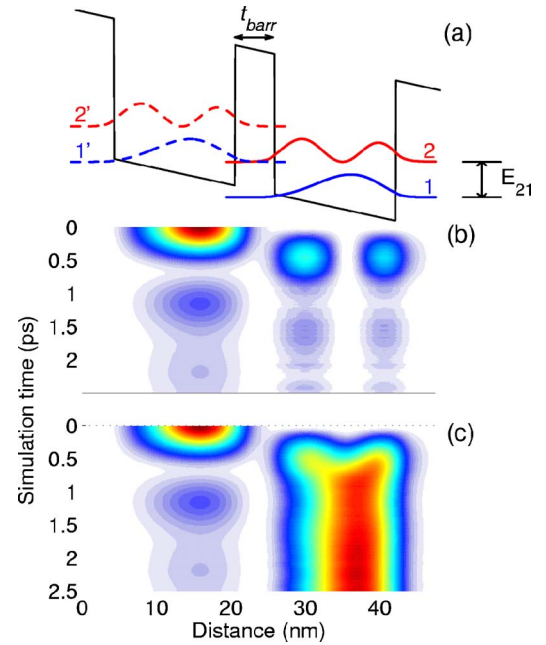

FIG. 2. Time evolution of the population density in a superlattice with 16.8-nm-wide GaAs wells and 4.1-nm-wide $Al_{0.15}Ga_{0.85}As$ barriers. The ground level in one well is in resonance with the first excited level in the next well ($\Delta_{1'2}=3.9$ meV). Level $n=2$ is depopulated by LO-phonon scattering ($\tau_{21}^{LO} \approx 0.3$ ps). A population density of 1×10^{10} cm^{-2} is initially placed in $n=1'$. (a) Conduction band profile and wavefunction probability distributions of the generic superlattice structure used throughout this paper. Also indicated are the barrier thickness t_{barr} and energy separation E_{21}, which are referenced throughout. (b) Spatial distribution of the populations of $n=1'$ and $n=2$ vs time. The damped Rabi oscillations of the population density are clearly visible. (c) Time evolution of the total population density in $n=1',2$, and the ground state 1.

module, and the wavefunctions are localized in the wells using the procedure outlined previously. The anticrossing gap between $n=1'$ and $n=2$ is given by $\Delta_{1'2}=2\Delta V_{1'2}$. The equations of motion can be written as[19]

$$\frac{d}{dt}(\rho_{11} - \rho_{22}) = \frac{2i\Delta V_{1'2}}{\hbar}(\rho_{21} - \rho_{21}^*)$$
$$- \frac{(\rho_{11} - \rho_{22}) - (\rho_{11} - \rho_{22})_0}{\tau_2}, \tag{27}$$

$$\frac{d}{dt}\rho_{21} = \frac{i\Delta V_{1'2}}{\hbar}(\rho_{11} - \rho_{22}) - \frac{iE_{1'2}}{\hbar} - \frac{\rho_{21}}{\tau_{deph}}, \tag{28}$$

where $(\rho_{11} - \rho_{22})_0$ is the population difference at equilibrium, and $\tau_{deph}^{-1} = 0.5\tau_2^{-1} + T_2^{-1}$. These coupled equations can be solved to find an expression for the current density through the barrier[16,17]:

$$J = qN_s \frac{|\Omega|^2 \tau_{deph}}{1 + (E_{1'2}/\hbar)^2 \tau_{deph}^2 + \Omega^2 \tau_2 \tau_{deph}}, \tag{29}$$

where $E_{1'2}$ is the energy detuning from resonance, $\Omega = 2\Delta V_{1'2}/\hbar$ is the Rabi oscillation frequency at resonance, and N_s is the total electron sheet density per well. This expression describes the current density versus detuning bias $E_{1'2}$ as a Lorentzian with a full width at half-maximum of
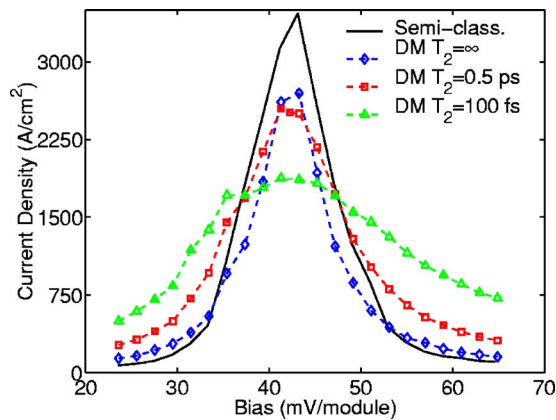
FIG. 3. Dependence of current density on bias for a superlattice consisting of 15.5-nm-wide GaAs wells, separated by 4.1-nm $Al_{0.15}Ga_{0.85}As$ barriers. The energy separation $E_{21} \approx 42$ meV and $\Delta_{1'2} = 4.5$ meV. Results are shown for both the semiclassical and the density-matrix simulations.

$$\Delta E_{FWHM} = \frac{2\hbar}{\tau_{deph}}[1 + \Omega^2 \tau_2 \tau_{deph}]^{1/2}. \tag{30}$$

The density matrix model provides a picture of a wave packet oscillating between $n=1'$ and $n=2$ at a frequency Ω with a damping rate of τ_{deph}^{-1}. This is illustrated in Figs. 2(b) and 2(c), which show the spatial distribution of the population density in a superlattice designed to study the dynamics of the depopulation of the lower radiative level 4 in a 3.2 THz QCL investigated later in this section [Fig. 5(b)]. The electron extraction from level 4 critically relies on the anticrossing with another level 3, which is subject to resonant LO-phonon scattering ($\tau_{LO} < 0.3$ ps). In our simple model we identify 4 and 3 (in the QCL) with the ground state 1' and the first excited state 2, respectively. The anticrossing energy $\Delta_{1'2} = 2\Delta V_{1'2}$ is 3.9 meV and E_{21} is 39 meV. A population of 1×10^{10} cm^{-2} is initially situated in 1' and then oscillates back and forth between the anticrossed levels 1' and 2. There is no pure dephasing added in this simulation ($1/T_2 = 0$). Electrons in state 2 are subject to resonant LO-phonon depopulation, which results in a damped Rabi oscillation with $\tau_{deph} \approx 0.6$ ps. From Fig. 2(c), we can see it takes about 1.5 ps for the majority of the electrons to transfer from 1' (corresponding to the lower radiative level 4) to 1, which corresponds to the relaxation/injection states in the QCL. This is slower than what we would expect from the semiclassical picture, where $\tau_4 \approx 0.5$ ps.[20] Note that even in the absence of dephasing, it would take approximately $h/2\Omega \approx 0.5$ ps for the electrons to oscillate across the barrier, which largely explains the longer dwell time in the density matrix calculation.

The dependence of the current density on the pure dephasing rate T_2^{-1} is shown in Fig. 3 for a similar superlattice. At resonance, the peak current density decreases with T_2^{-1} as the Rabi oscillation is increasingly damped and it becomes more and more difficult for electrons to tunnel through the barrier. The difference between the semiclassical result and the density matrix simulation in the limit of $T_2^{-1} = 0$ is due to the relaxation dephasing, which is the term $-1/2(\gamma_i + \gamma_j)$ in Eq. (13). Away from resonance, the DM model predicts a broader I-V curve with a higher current
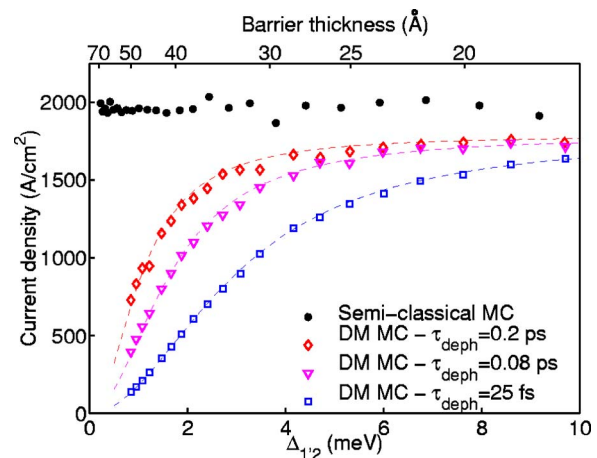

FIG. 4. Dependence of peak current density on anticrossing gap $\Delta_{1'2}$ for a GaAs/$Al_{0.3}Ga_{0.7}As$ superlattice with 14.8-nm-thick wells separated by barriers with varying thicknesses. The energy separation $E_{21} \approx 60$ meV. The semiclassical results are independent of $\Delta_{1'2}$ and the barrier thickness, whereas the density matrix simulation clearly shows the decrease of the peak current density with smaller $\Delta_{1'2}$ and higher dephasing rate T_2^{-1}. The dashed lines represent theoretical calculations using Eq. (29).

density than the semiclassical picture. This is due to the level broadening [Eq. (30)] which relaxes the energy alignment of $n=1'$ and $n=2$. Especially at high dephasing rates, the increased interaction can cause the upper radiative state (not shown here) to couple more strongly to the injector/relaxation level 1, and thus reduce the depopulation selectivity.

The influence of the barrier thickness t_{barr} is felt through a diminished coupling $\Delta V_{1'2}$, and can be illustrated by investigating the current transport through a barrier. We consider a superlattice structure similar to the one shown in Fig. 2(a), with an energy separation $E_{21} \approx 60$ meV which is much larger than $\hbar\omega_{LO}$. This ensures that, in the semiclassical picture, the intrawell energy difference E_{21} still exceeds $\hbar\omega_{LO}$ even for large anticrossing gaps so that scattering from $n=2$ is dominated by LO-phonon scattering ($\tau_2^{LO} \approx 0.3$ ps, and nearly independent of E_{21}). The peak current density occurs when the lower level 1' of one well lines up with the upper level 2 of the adjacent well. In the semiclassical approach, the anticrossed wavefunctions are delocalized across both wells, and share an identical carrier lifetime τ (independent of barrier thickness) disregarding the minor energy shift due to the anticrossing. This leads to the unphysical result that the peak current density in the semiclassical description does not depend on the barrier thickness (as shown in Fig. 4). In other words, in this picture electrons scatter from one spatially extended state into the next, and never experience any effect from the barrier whatsoever, as illustrated previously in Fig. 1. To include tunneling effects, and hence more accurately describe transport through a barrier, we need to take into account the phase correlation between localized basis states, so that phase relaxation results in a collapse into localized states, thus interrupting resonant tunneling.

The results for the density matrix calculations are also shown in Fig. 4. In contrast to the semiclassical results, the DM-MC calculations reveal a strong dependence of the peak current density on the anticrossing gap (or barrier thickness)

and the pure dephasing time T_2. For large anticrossing gaps or thinner barriers, the Rabi oscillation frequency is much higher than the dephasing rate and the peak current density approaches the semiclassical limit. In this regime the population is spread equally across both subbands, and the current density is given approximately by $J = qN_s/2\tau_2$. For smaller anticrossing gaps, or thicker barriers, the dephasing scattering becomes increasingly important and inhibits the transport through the barrier, which results in a lower peak current density. From Eq. (29) we can see that the current density starts to roll off when $\Delta_{1'2} \approx \hbar/\sqrt{\tau_2\tau_{\text{deph}}}$. The dashed lines are analytical results calculated from Eq. (29), which agree well with our DM-MC results. The good agreement based on this simple structure gives us confidence to investigate more complicated structures using the numerical DM-MC tool developed in this work.

Note that localized basis wavefunctions are necessary to produce the above result within the density matrix formalism. If we chose spatially extended basis wavefunctions (as in the semiclassical case), a localized wave packet would be represented as a coherent superposition of extended wavefunctions:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|S\rangle + |A\rangle) \to \rho = |\psi\rangle\langle\psi|. \qquad (31)$$

In the absence of dephasing (and scattering), the time evolution of this wave packet would be identical to the one described with localized basis wavefunctions. A proper implementation of scattering requires a full density matrix approach,[21] which is not expected to result in a substantial difference between the two models. However, the inclusion of pure dephasing causes the off-diagonal elements of ρ to decrease, so that even in the strong dephasing limit of $T_2 \to 0$:

$$\rho = \frac{1}{2}|S\rangle\langle S| + \frac{1}{2}|A\rangle\langle A|, \qquad (32)$$

which is still equivalent to the semiclassical model, and no decrease in current density is predicted. This means that extended wavefunctions are not a good choice for the basis wavefunctions in our model.

To investigate the utility of the density matrix method in the modeling of electron transport in QCLs, the density matrix model and a semiclassical MC simulation were used to calculate current densities, populations and electron temperatures for several different QC designs. Here we focus on two representative examples, for which experimental results were published. As mentioned in Sec. III, the simulations include semiclassical electron-phonon (acoustic and LO), electron-impurity (e-imp) and electron-electron (e-e) scattering, and a nonequilibrium, multisubband screening model for e-imp and e-e interactions. The only phenomenological parameter is the pure dephasing time T_2, used in the DM-MC calculations. In all simulations, except for Fig. 8, the lattice temperature T_{latt} was assumed to be 25 K.

The first investigated device is a simple double-quantum-well structure[7] (T65), whose conduction band diagram and wavefunctions are reproduced here in Fig. 5(a). In
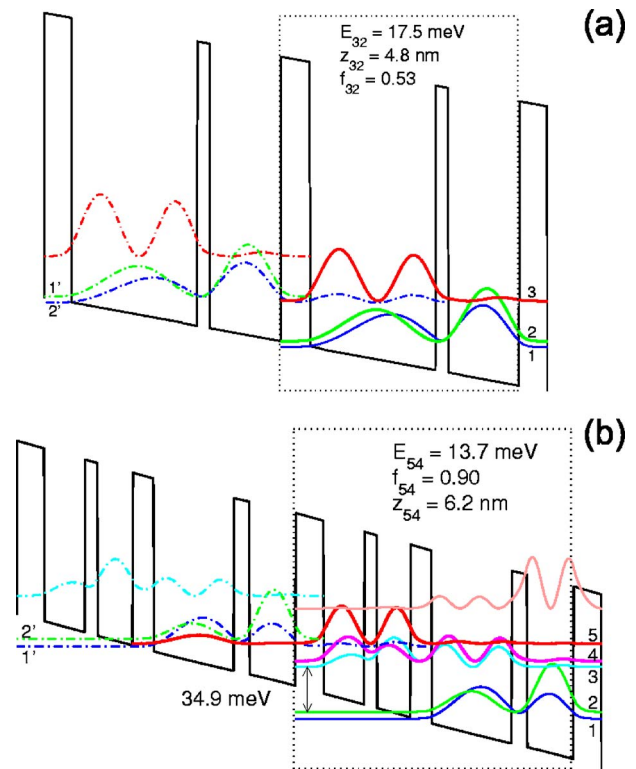

FIG. 5. (a) Band structure for T65. The device consists of GaAs/Al$_{0.15}$Ga$_{0.85}$As layers with thicknesses (nm) 5.5/23.4/2.4/13.2 (barriers in boldface, wells in plain text) and is doped to $n = 1.4 \times 10^{16}$ cm^{-3} in the 13.2-nm-wide well, resulting in a sheet density of 1.85×10^{10} cm^{-2} per module. (b) Band structure for FL177C-M5, which consists of GaAs/Al$_{0.15}$Ga$_{0.85}$As layers with thicknesses (nm) 5.5/7.9/2.5/6.5/4.1/15.5/3.0/9.0 and is doped to $n = 1.9 \times 10^{16}$ cm^{-3} in the 15.5-nm-wide well, resulting in a sheet density of 3×10^{10} cm^{-2} per module.

experiments, the current characteristics of this device were nearly independent of temperature for a lattice temperature between 5 and 77 K, whereas the electroluminescence (proportional to τ_3) dropped by a factor of 1.5. No gain or superlinear power-current (L-I) relations were ever observed. This is a strong indication that transport through the thick injector barrier is limited by incoherent tunneling between the injector state $n=1'$ and the upper radiative level $n=3$ ($\Delta_{1'3} \approx 0.8$ meV). Furthermore, magnetotunneling spectroscopy revealed clear evidence of an anticrossing gap of $\Delta_{21} \approx 2$ meV between levels 1 and 2, validating our hybrid model of using a semiclassical model for intramodule transport. Table I and Fig. 6 present the main calculation results with and without coherent transport. The DM simulations included a phenomenological dephasing time of 0.5 ps,

TABLE I. Calculated subband energy, temperature, and population density of T65 at injection anticrossing ($T_{\text{latt}}=25$ K) for the semiclassical and density matrix simulations.

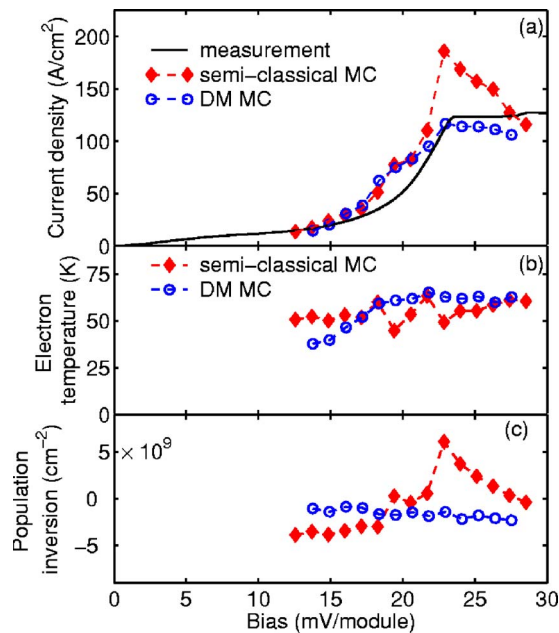|		T_{el} (K)		Population (10^{10} cm^{-2})	
n	E (meV)	Semiclassical	DM	Semiclassical	DM
1	0	50	86	0.53	0.91
2	3.9	85	95	0.37	0.58
3	25.0	64	62	0.95	0.42

FIG. 6. Key results of the MC simulations for T65. The lattice temperature is assumed to be 25 K. (a) Current density for a range of biases. The density matrix results were obtained with a phenomenological pure dephasing time T_2=0.5 ps. (b) Electron temperature for the upper subband involved in the radiative transition, n=3. (c) The calculated population inversion $n_3 - n_2$, which is proportional to the gain.

which is consistent with the measured spontaneous emission linewidth Δf of 2 meV (0.5 THz, and $\Delta f \approx h/\pi T_2$). The DM calculations show a peak current density of 116 A/cm^{-2} that corresponds well with the experiments (123 A/cm^{-2}), whereas the semiclassical model overestimated the peak current density (186 A/cm^{-2}). In both simulations there is a bump in the I-V characteristic around 20 mV/module, close to the anticrossing of n=2' and n=3. However, there is no evidence of this transition in the experimental results. The sharp features in both the current density and electron temperature, which are clearly visible in the semiclassical calculation, are smoothed out in the DM results. Most important is the absence of population inversion, and hence gain, in the DM calculation, which is in agreement with the experimental results. On the other hand, the semiclassical results indicate a maximum population inversion of $n_3 - n_2 \approx 6 \times 10^9$ cm^{-2} (corresponding to a $\Delta n_{3D} \approx 1.4 \times 10^{15}$ cm^{-3}), and a predicted gain of 180 cm^{-1} for a spontaneous emission linewidth Δf =2 meV. Such a high level of gain would have made lasing quite easy to achieve. However, no lasing was observed from T65 and similar structures, even embedded in metal-metal waveguides.[19] Our DM analysis indicates that it is likely that electrons are "stuck" behind the injection barriers, residing mostly in levels 1' and 2' instead of equally populating 1' and 3 as the semiclassical model predicted.

The second QC device is a 3.2 THz QCL,[8] labeled FL177C-M5, which operated up to ~130 K in pulsed mode, and whose band diagram and wavefunctions are reproduced in Fig. 5(b). Table II and Fig. 7 show a comparison of the key MC results with and without coherent transport (T_2 =0.33 ps, to reproduce a FWHM linewidth of 6 meV together with relaxation scattering $\tau_4 \approx 0.5$ ps). From Fig. 7(a),

TABLE II. Calculated subband energy, temperature, and population density of the 3.2 THz laser at the peak current density predicted by the DM simulation (V_{bias}=59 mV/module), for the semiclassical and density matrix simulations. The lattice temperature is 25 K.

		T_{el} (K)		Population (10^{10} cm^{-2})	
n	E (meV)	Semiclassical	DM	Semiclassical	DM
1	0	114	143	1.14	1.27
2	5.4	127	138	0.74	0.93
3	41.3	152	174	0.13	0.10
4	46.3	154	174	0.19	0.11
5	60.4	115	108	0.73	0.52

we can see that the semiclassical results exhibit a large parasitic current at biases around 40 mV/module, corresponding to the 1' → 3 (with $\Delta_{1'3}$=0.45 meV) transition. Similar to the 3.4 THz QCL explored in Ref. 4, the overestimation of the current density in this parasitic channel is due to the use of extended basis states in the calculation of the scattering rates. The density matrix approach largely eliminates this problem, but there is still a noticeable hump in the I-V whereas there is none in the measurements, although the presence of the parasitic channel is still evident in differential conductance measurements.[8] This simulation result corresponds well with the full quantum mechanical calculations done for this same structure in Ref. 22, which is a further indication that the used DM approach adequately models the coherence effects in the electron transport. The predicted DM peak current density (700 A/cm^{-2}) and gain are achieved when the upper radiative level is lined up between the 2'-5 and 1'-5 anti-
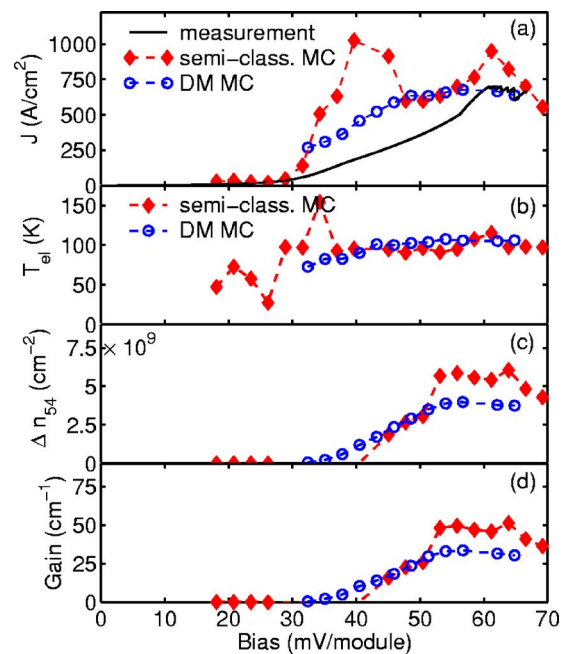

FIG. 7. Key results of the MC simulations for the 3.2 THz QCL. The lattice temperature is assumed to be 25 K. (a) Current density for a range of biases. The density matrix results were obtained with a phenomenological pure dephasing time of 0.33 ps. (b) Electron temperature for the upper subband involved in the radiative transition, n=5. (c) The calculated population inversion $n_5 - n_4$. (d) Calculated gain for a spontaneous emission linewidth of 6 meV.
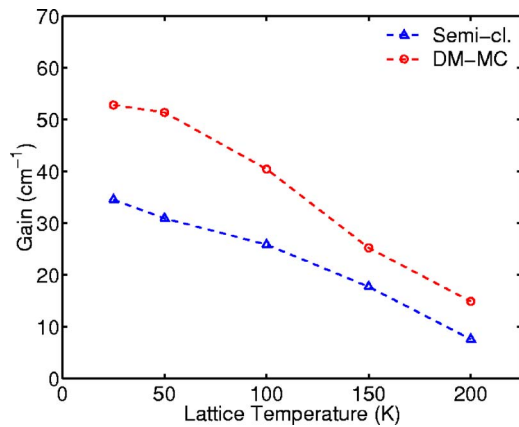
FIG. 8. Maximum gain of FL177C-M5 as a function of lattice temperature for the semiclassical and DM-MC simulations, assuming a spontaneous emission linewidth of 6 meV. In experiments, the lattice temperature is somewhat higher than the heat-sink temperature, and the devices lase up to ~130 K heat-sink temperature.

crossings, and not at the $1'$-5 anticrossing as for the semiclassical simulation. Due to the finite linewidth of the levels [>4 meV, see Eq. (30)], which is comparable to $E_{2'1'}$ ≈ 5 meV, their similar anticrossing gaps ($\Delta_{1'5} \approx \Delta_{2'5}$ ~ 1.8 meV) and the similar population densities of $1'$ and $2'$, injection is most efficient when both $1'$ and $2'$ contribute. Note that for small linewidths, injection from $2'$ and $1'$ into 5 is more centered around their respective anticrossing bias, which may result in a current peak at the $2'$-5 anticrossing, followed by a region of negative differential resistance (NDR). In experiments, this early NDR would prevent the device from reaching its designed level alignment and degrade its performance. The predicted peak population inversion ΔN_{54} and gain g are lower for the DM simulations than for the semiclassical calculations, $\Delta N_{54} = 4.1 \times 10^9$ cm^{-2} (g =32 cm^{-1} for Δf=6 meV) and $\Delta N_{54}=6.4 \times 10^9$ cm^{-2} (g =52 cm^{-1}), respectively. This is largely due to a decreased injection efficiency and selectivity, as explained in the discussion of Fig. 3. To investigate the high-temperature performance of the QCL, simulations for lattice temperatures up to 200 K were performed. As can be seen from Fig. 8, both simulations predict a steady decrease of the gain to g =8 cm^{-1} (DM) and 14 cm^{-1} (semiclassical) at 200 K. This decline is mainly due to the increased LO-phonon mediated depopulation of level 5. In experiments, CW lasing was observed in very small devices (100×100 μm^2) immersed in liquid nitrogen, which roughly corresponds to a facet mirror loss of 40 cm^{-1}, for a facet reflectivity of ~60%.[23] This experimentally inferred gain agrees reasonably well with the predicted gain of ~30 cm^{-1} (DM) and 44 cm^{-1} (semiclassical) at 77 K. The underestimation of gain and peak current density in the DM simulation indicates that the used pure dephasing time T_2=0.33 ps may be too short.

Even though the difference between the peak gain and the current densities calculated from the DM-MC and semiclassical MC is only quantitative, there is an important qualitative difference between the two calculations in the current-voltage (*I-V*) characteristics. As can be seen in Fig. 7(a), the semiclassical current density at the $1'$-3 parasitic channel is higher than that at the designed bias of the $1'$-5 anticrossing.

Experimentally, this would have resulted in a negative differential resistance (NDR) above the bias at the $1'$-3 anticrossing, making the desired $1'$-5 energy alignment inaccessible. Fortunately, dephasing reduces the current density at this parasitic channel much more than at the designed bias due to the smaller anticrossing gap $\Delta_{1'3} \approx 0.5$ meV, compared to $\Delta_{1'5} \approx 1.8$ meV. Consequently, we were able to bias the 3.2 THz laser device at the $1'$-5 energy alignment and achieve lasing. However, for lower-frequency QCLs at 2.1 THz, the biases of $1'$-3 and $1'$-5 are closer, resulting in much more similar current densities at those respective bias conditions and a much smaller range of current densities in which the devices laser.[24] In this important aspect, the semiclassical MC calculation is useless in predicting the relative current densities at the $1'$-3 and $1'$-5 biases, and whether a NDR will occur. With the much smoother *I-V* that is closer to the experimental results, we hope that the DM-MC could help us in designing suitable structures with a reduced $1'$-3 parasitic current density for lower-frequency lasers.

V. CONCLUSION

In conclusion, as a first step we have shown that the inclusion of a model for coherent transport and dephasing is essential to describe the transport dynamics of intersubband transport in THz QCLs. The density matrix model, together with the choice of localized basis states, allows for an intuitive treatment of transport between weakly coupled levels and the incorporation of the effects of sequential tunneling into a MC simulation. We have used a semiclassical and a density matrix MC simulation to compare calculated current densities and gain with experimental measurements. The inclusion of coherent transport showed marked improvement over the semiclassical model. It largely eliminated the overestimation of the peak current density and parasitic current channels, and correctly predicted the absence of a population inversion where the semiclassical model predicted a large gain. However, more remains to be done. The use of a single, phenomenological pure dephasing time to describe the interaction between all subbands does not take into account the substantial differences in elastic intrasubband (impurity and interface roughness) scattering for different levels in a module. The different models used to describe these scattering mechanisms and the large scattering rates these models predict for small-angle scattering, make it difficult to dependably estimate their contribution to dephasing. More detailed calculations incorporating this subband-dependent dephasing can provide a model that uses no phenomenological parameters, and will yield a more accurate description of the electron transport. This more comprehensive simulation can be a valuable tool for designing and analyzing QCLs.

ACKNOWLEDGMENTS

[1] H. Vaupel, P. Thomas, O. Kuehn, V. May, K. Maschke, A. P. Heberle, W. W. Ruehle, and R. Koehler, Phys. Rev. B **53**, 16531 (1996).

[2] F. Eickemeyer, K. Reimann, M. Woerner, T. Elsaesser, S. Barbieri, C. Sirtori, G. Strasser, T. Mueller, R. Bratschitsch, and K. Unterrainer, Phys. Rev. Lett. **89**, 047402 (2002).

[3] C. W. Luo, K. Reimann, M. Woerner, T. Elsaesser, R. Hey, and K. H. Ploog, Semicond. Sci. Technol. **19**, S285 (2004).

[4] H. Callebaut, S. Kumar, B. S. Williams, Q. Hu, and J. L. Reno, Appl. Phys. Lett. **84**, 645 (2004).

[5] S. C. Lee and A. Wacker, Appl. Phys. Lett. **83**, 2506 (2003).

[6] H. Willenberg, G. Doehler, and J. Faist, Phys. Rev. B **67**, 085315 (2003).

[7] B. S. Williams, H. Callebaut, Q. Hu, and J. L. Reno, Appl. Phys. Lett. **79**, 4444 (2001).

[8] S. Kumar, B. S. Williams, S. Kohen, Q. Hu, and J. L. Reno, Appl. Phys. Lett. **84**, 2494 (2004).

[9] S. Luryi, Appl. Phys. Lett. **47**, 490 (1985).

[10] H. Callebaut, S. Kumar, B. Williams, Q. Hu, and J. Reno, Appl. Phys. Lett. **83**, 207 (2003).

[11] A. Mircetic, D. Indjin, Z. Ikonic, P. Harrison, V. Milanovic, and R. Kelsall, J. Appl. Phys. **97**, 84506 (2005).

[12] R. Koehler, R. C. Iotti, A. Tredicucci, and F. Rossi, Appl. Phys. Lett. **78**, 2904 (2001).

[13] B. Williams, S. Kumar, H. Callebaut, Q. Hu, and J. Reno, Appl. Phys. Lett. **83**, 5142 (2003).

[14] S. Mukamel, in *Principles of Nonlinear Optical Spectroscopy* (Oxford University Press, New York, 1995).

[15] R. C. Iotti and F. Rossi, Phys. Rev. Lett. **87**, 146603 (2001).

[16] R. F. Kazarinov and R. A. Suris, Sov. Phys. Semicond. **5**, 707 (1971).

[17] C. Sirtori, F. Capasso, J. Faist, A. L. Hutchinson, D. L. Sivco, and A. Y. Cho, IEEE J. Quantum Electron. **34**, 1722 (1998).

[18] S. M. Goodnick and P. Lugli, Appl. Phys. Lett. **51**, 584 (1987).

[19] B. S. Williams, Ph.D., thesis, M.I.T., Department of Electrical Engineering and Computer Science (2003).

[20] B. Williams, H. Callebaut, S. Kumar, Q. Hu, and J. Reno, Appl. Phys. Lett. **82**, 1015 (2003).

[21] T. Kuhn and F. Rossi, Phys. Rev. B **46**, 7496 (1992).

[22] F. Banit, S. C. Lee, A. Knorr, and A. Wacker, Appl. Phys. Lett. **86**, 041108 (2005).

[23] S. Kohen, B. Williams, and Q. Hu, J. Appl. Phys. **97**, 053106 (2005).

[24] B. Williams, S. Kumar, Q. Hu, and J. Reno, Electron. Lett. **40**, 431 (2004).