# Coherent transport in semiconductor heterostructures: A phenomenological approach

Ariel Gordon* and Daniel Majer
*Optigo Systems Ltd., 1 Hamelacha Street, Lod 71520, Israel*

(Received 9 September 2009; revised manuscript received 21 October 2009; published 18 November 2009)

We propose a theoretical method for describing coherent quantum transport in semiconductor heterostructures and particularly in quantum cascade lasers (QCLs). The method is an extension of standard rate-equation models to include coherence. Instead of building the model from microscopic considerations, we construct it from the following requirements: invariance under basis change, non-negativity of the density matrix, and compatibility with existing rate-equation models. The computational effort associated with the proposed method is very modest. It is shown that the role of coherence in QCLs is crucial and omitting it from the calculation leads to unphysical results. The discussion in the paper is focused on QCLs but the approach applies in general to semiconductor heterostructures.

DOI: 10.1103/PhysRevB.80.195317 PACS number(s): 73.40.Kp, 42.55.Px, 73.63.Hs

## I. INTRODUCTION

The first demonstration of the quantum cascade laser (QCL) (Ref. 1) brought up the need for theoretical understanding of the underlying physics. One of the most intriguing theoretical questions about QCLs concerns the nature of current transport in these devices. While in many works it is argued that transport is mainly incoherent,[2,3] others emphasize the crucial role played by coherence QCLs.[4–6]

Including coherence in a QCL model is challenging. In Ref. 4 phenomenological dephasing terms are manually introduced into the equations of a few chosen subbands, which are then treated differently from the others, similarly to the Kazarinov-Suris tight-binding approach.[7] The only comprehensive theoretical approach known to us that includes coherence is the nonequilibrium Green's-function method.[5] This approach is computationally extremely demanding.

We propose an approach to the issue of coherence in QCLs, which leads to a simple and comprehensive model. The computational effort associated with the model is similar to that of incoherent rate-equation models,[8] which is very modest. The voltage and emitted optical power as function of the current are calculated within minutes on a laptop computer. The results well compare to published experimental results.

Our approach is inspired by the Landau theory of phase transitions.[9] Landau used symmetries of physical systems to propose forms for their free energies. Once invariance under these symmetries is imposed, the "freedom of choice" of the model rests in relatively few parameters. In Landau's approach it is unnecessary to understand in detail the microscopy of the system. Whatever the microscopy may be, it eventually must lead to a model that respects the symmetries of the system.

We employ the same concept in order to obtain a phenomenological model of a QCL that includes the effects of coherence. The invariance in this case is with respect to the choice of the quantum-mechanical basis functions: any physical model must give results (for measurable quantities) that are independent of the choice of basis. To this principle we add two constraints: first is non-negativity: the density matrix must be non-negative definite; second is compatibility: the dynamics of populations should agree with existing (and very well-studied) rate-equation models.[8] The information on coherences in QCLs is therefore already there, hidden in the existing incoherent models.

## II. BASIS INVARIANCE

The starting point of our model is an equation of motion for the density matrix for the populations and coherences of *entire subbands*

$$\rho_{fg} = \frac{i}{\hbar}[\rho, H]_{fg} + \sum_{ij} R_{ijfg}\rho_{ij}. \tag{1}$$

$\rho$ is the density matrix integrated over the lateral degrees of freedom $(x,y)$, and is thus suitable for evaluating expectation values of operators independent of $x$ and $y$, such as the current in the growth direction $z$. $\rho_{ii}$ is proportional to the number of electrons in the $i$th subband, whereas $\rho_{i\neq j}$ is the coherence between electrons in subbands $i$ and $j$, averaged over the lateral degrees of freedom. $H$ is the unperturbed Hamiltonian of the heterostructure in the growth direction $z$. The $R_{ijfg}$ coefficients represent all the physical processes included in the model that are not represented in $H$. The subscripts $i$, $j$, $f$, and $g$ refer at this stage to some general set of basis functions.

For the expectation values not to depend on the choice of basis, $\rho$ must transform as

$$\tilde{\rho}_{ij} = \sum_{i',j'} U_{ii'}U^*_{jj'}\rho_{i'j'}, \tag{2}$$

where $U_{ii'}$ is a unitary matrix representing the basis change and $\tilde{\rho}_{ij}$ is the density matrix represented in the new basis. From Eq. (2) it follows that $R_{ijfg}$ must transform under the change of basis as

$$\tilde{R}_{ijfg} = \sum_{i',j',f',g'} U_{ff'}U^*_{gg'}U^*_{ii'}U_{jj'}R_{i'j'f'g'}. \tag{3}$$

We now argue that Eq. (3), along with the known diagonal scattering rates $R_{iiff}$, determines to large extent the general $R_{ijfg}$ rates and thereby the dynamics of coherences.

 

For the clarity of presentation we demonstrate our approach on one mechanism only, namely, longitudinal-optical (LO) phonons. LO phonons are a key mechanism in QCLs,[1] and they will serve the reader as an example of the general procedure. For the same reason we make a few simplifying assumptions: we assume that the population in all subbands is: (1) equilibrated and (2) Maxwell-Boltzmann distributed. The justification of (2) lies in the low-doping densities in QCLs, which are far below degeneracy at least at noncryogenic temperatures. Assumption (1) is justified by the fast equilibration of the electron distribution within a subband.[10] Deviations from equilibrium subband distributions have been observed[11] but we neglect them for the time being. Additionally, we assume that all subbands have the same electron temperature $T_e$.[10] The above assumptions are mainly made for simplicity and are not necessary for invoking the invariance principle.

We assume the common Fröhlich description of the electron-LO-phonon interaction.[12] Equations (13) and (14) of Ref. 12 give the phonon emission rate associated with an intersubband transition, assuming a particular initial electron kinetic energy. We average these expressions from Ref. 12 over a Maxwell-Boltzmann distribution of initial kinetic energies. After a lengthy but straightforward calculation we obtain the following transition rate

$$\Gamma_{if} = \int_{-\infty}^{\infty} |A_{if}(k)|^2 g_{if}(k) dk. \tag{4}$$

The subscripts $i$ and $f$ denote the initial and final subband, which are eigenstates of $H$, and

$$A_{if}(k) \equiv \int_{-\infty}^{\infty} \psi_f^*(z)\psi_i(z)e^{-ikz}dz, \tag{5}$$

$$g_{if}(k) \equiv C\int_0^{\infty} \frac{dq}{q^2+k^2}\exp\left(-\frac{[\hbar^2 q^2 - 2m^*\Delta_{if}]^2}{8m^* k_B T_e \hbar^2 q^2}\right), \tag{6}$$

$$\Delta_{if} \equiv \varepsilon_i - \varepsilon_f - \hbar\omega_{\mathrm{LO}}. \tag{7}$$

$C = \frac{\omega_{\mathrm{LO}}(n+1)e^2}{8\pi\hbar\epsilon_0\epsilon_p}\sqrt{\frac{m^*}{2\pi k_B T_e}}$, where $\hbar\omega_{\mathrm{LO}}$, $n$, and $m^*$ are the optical-phonon energy, the thermal phonon population, and the effective mass, respectively. $\epsilon_p \equiv (\epsilon_\infty^{-1} - \epsilon_s^{-1})^{-1}$, where $\epsilon_\infty$ and $\epsilon_s$ are the optical and static dielectric constants. $\varepsilon$ and $\psi$ denote subband energies and wave functions, respectively. Equation (4) gives immediately the rates in Eq. (1) connecting populations, namely, $R_{iiff}$,

$$R_{iiff} = \Gamma_{if} - \delta_{if}\sum_m \Gamma_{im}. \tag{8}$$

We now turn to the task of inferring a general $R_{ijfg}$ coefficient from Eq. (8). Since the derivation is lengthy, we show here only its first steps and the final result, whereas the intermediate steps are given in Appendix A. We start with $R_{iifg}$, where $i \neq f$, $i \neq g$, and $f \neq g$. We first consider the limit where the subbands $|f\rangle$ and $|g\rangle$ approach degeneracy. In this case we are entitled to change the basis from $|f\rangle$ and $|g\rangle$ to

$$|+\rangle \equiv \frac{|f\rangle + |g\rangle}{\sqrt{2}}, \quad |-\rangle \equiv \frac{|f\rangle - |g\rangle}{\sqrt{2}}, \tag{9}$$

and to calculate the scattering rates $R_{ii++}$ and $R_{ii--}$ using Eq. (4). From Eq. (3) we obtain the relation

$$R_{iifg} + R_{iigf} = \Gamma_{i+} - \Gamma_{i-}. \tag{10}$$

The same procedure can be repeated with the transformation

$$|+'\rangle \equiv \frac{|f\rangle + i|g\rangle}{\sqrt{2}}, \quad |-'\rangle \equiv \frac{|f\rangle - i|g\rangle}{\sqrt{2}}, \tag{11}$$

leading to

$$i(R_{iifg} - R_{iigf}) = \Gamma_{i+'} - \Gamma_{i-'}. \tag{12}$$

For a QCL structure one can always choose all subband eigenfunctions to be real.[13] Then one can see from Eqs. (4)–(7) that the right-hand side of Eq. (12) is zero. We thus find that

$$R_{iifg} = \frac{\Gamma_{i+} - \Gamma_{i-}}{2} = \int_{-\infty}^{\infty} A_{if}(k)A_{ig}^*(k)g_{if}(k)dk. \tag{13}$$

Following the same steps that led to Eq. (13), with other basis transformations (see Appendix A), one can arrive at a general expression for $R_{ijfg}$,

$$R_{ijfg} = Q_{ijfg}^0 - \frac{1}{2}\delta_{gj}\sum_m Q_{ifmm}^0 - \frac{1}{2}\delta_{if}\sum_m Q_{gjmm}^0, \tag{14}$$

where

$$Q_{ijfg}^0 \equiv \int_{-\infty}^{\infty} A_{if}(k)A_{jg}^*(k)g_{if}(k)dk. \tag{15}$$

## III. NON-NEGATIVITY

The calculation so far holds only at the limit where the subbands are exactly degenerate. For example, Eq. (13) holds provided that $\varepsilon_f = \varepsilon_g$. The degeneracy was explicitly made use of: Eq. (4) relies on perturbation theory, which assumes that all the $\psi$-$s$ represent eigenstates of the unperturbed Hamiltonian. The latter is not true for $|\pm\rangle$ and $|\pm'\rangle$ if the subbands $f$ and $g$ are not degenerate.

It turns out however that coherence is non-negligible only *near* a degeneracy. To show that we consider two nondegenerate subbands $f$ and $g$. Equation (1) written in the eigenbasis of $H$ reads

$$\dot{\rho}_{fg} = \frac{i}{\hbar}\rho_{fg}(\varepsilon_g - \varepsilon_f) + R_{fgfg}\rho_{fg} + D_{fg}, \tag{16}$$

where

$$D_{fg} \equiv \sum_{(i,j)\neq(f,g)} R_{ijfg}\rho_{ij}. \tag{17}$$

In the equation of motion or $\rho_{fg}$ [Eq. (16)], the dephasing coefficient $R_{fgfg}$, which is negative, acts as damping, whereas the coupling to the other coherences and populations ($D_{fg}$)

acts as driving. Consider for the moment only driving terms with $i=j$, i.e., populations.[14] Then we can assume that $D_{fg}$ varies very slowly over the time scale defined by $R_{fgfg}$. Under these conditions the order of magnitude of $\rho_{fg}$ is given by

$$\rho_{ij} \sim \frac{D_{fg}}{R_{fgfg} + i(\varepsilon_g - \varepsilon_f)}. \tag{18}$$

We can see that when $|\varepsilon_g - \varepsilon_f| \gg R_{fgfg}$, $|\rho_{ij}|$ is small. Therefore for dephasing times of 50 fs [which are on the short side in QCLs (Ref. 1)], it follows that $\rho_{fg}$ vanishes for $|\varepsilon_g - \varepsilon_f| \gg 13$ meV.

The first conclusion of the above discussion is that Sec. II brings us very close to our goal since it is only near degeneracy where coherences are required. The second conclusion is that the computational workload associated with implementing our model is comparable to that of incoherent rate equations because relatively few coherences are needed to be included. In practice, one can define a threshold for energy separation above which coherences are omitted and increase it until numerical convergence is reached.

We therefore only need to extend Eq. (13) from exact degeneracy to near degeneracy. Such an extension must comply with the requirement that Eq. (1) preserves the nonnegativity (and the trace) of $\rho$. This requirement alone strongly restricts the $R_{ijfg}$ coefficients. In fact, $\rho$ will remain non-negative if and only if $R_{ijfg}$ can be written in a very particular way, namely, in the Lindblad form.[15,16]

We propose the following generalization of Eqs. (14) and (15)

$$R_{ijfg} = Q_{ijfg} - \frac{1}{2}\sum_m Q_{ifmm}\delta_{gj} - \frac{1}{2}\sum_m Q_{gjmm}\delta_{if}, \tag{19}$$

where

$$Q_{ijfg} = \int_{-\infty}^{\infty} A_{if}(k)A_{jg}^*(k)\sqrt{g_{if}(k)g_{jg}(k)}dk. \tag{20}$$

In Appendix B it is shown that Eqs. (19) and (20) conform to the Lindblad form, which guarantees the preservation of nonnegativity and the trace of $\rho$ under the evolution described by Eq. (1).

So far we have shown that Eqs. (19) and (20) satisfy the requirements of basis invariance, non-negativity of the density matrix, and compatibility with incoherent rate-equation models. Therefore they form *a* legitimate model but there may be other choices that satisfy all the requirements listed above. Now we argue that the freedom to move away from Eqs. (19) and (20) is very limited.

Two choices were made in Eq. (20): for the wave-function-dependent part $A_{if}(k)A_{jg}^*(k)$ and for the energy-dependent part $\sqrt{g_{if}(k)g_{jg}(k)}$. For the former, we have chosen the same dependence as in the degenerate case Eq. (15). This choice can be motivated by the Kazarinov-Suris theory of resonant tunneling.[7] In their theory, the wave functions remain unchanged as the bias field is swept through the point of anticrossing. Leaving the wave-function-dependent part of Eq. (15) unchanged has a similar effect. Loosely speaking, the role of the wave-function-dependent part is to "undo" the mixing of the basis functions caused by tunneling, and thus to leave the scattering rates unaffected by the change in the underlying basis functions. By building a simple example with three wave functions, it is straightforward to show that leaving the wave-function-dependent part of Eq. (13) unchanged is needed for our model to agree with the Kazarinov-Suris theory. Once the wave-function-dependent part of Eq. (20) is fixed, the energy-dependent part is bound to be given by $\sqrt{g_{if}(k)g_{jg}(k)}$, up to a correction of second (or higher) order in the deviation from degeneracy (see Appendix B).

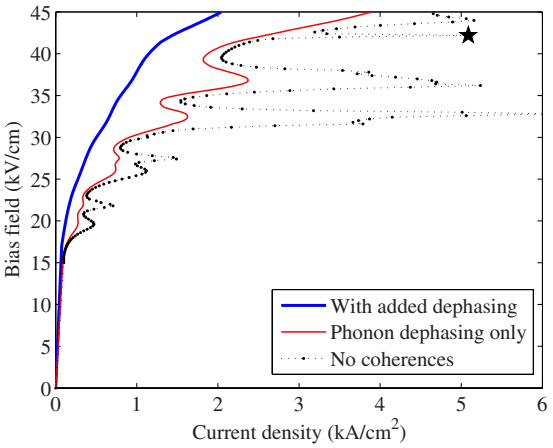


FIG. 1. (Color online) Calculated current density as function of the bias field. The black dotted, red thin continuous, and blue thick curves represent calculations without coherences at all, with coherences, and with coherences and added dephasing, respectively. The star denotes the peak which is analyzed in detail in Figs. 2 and 3.

Equation (19) gives the $R_{ijfg}$ coefficients associated with phonon emission. To construct a model of a QCL, one needs to include other nonradiative mechanisms[8] and radiative transitions. The former can be derived from standard well-accepted rate equations, repeating the procedure outlined above in Secs. II and III. Radiative $R_{ijfg}$ coefficients can be derived directly from the dipole interaction Hamiltonian, as a straightforward generalization of the Bloch equations[17] to a many-level system. We shall not pursue these derivations here but rather dedicate the remaining part of the paper to demonstrating some results and capabilities of the model.

## IV. COMPUTATIONAL EXAMPLE

For the demonstration we chose the design given in Ref. 18. The material parameters were taken from Ref. 1. Figure 1 shows the bias field and output power as function of the current density flowing through the laser. The lattice temperature was set to 300 K and $T_e$ was set to 300 K as well merely for simplicity. Without coherences, the current exhibits many sharp peaks as the bias field is varied. Each peak is associated with an anticrossing between wave functions. For example, the peak denoted by a star in Fig. 1 is associated with the anticrossing shown in Fig. 2. One can see that there are two wave functions extending from the "left" (see Fig. 2)

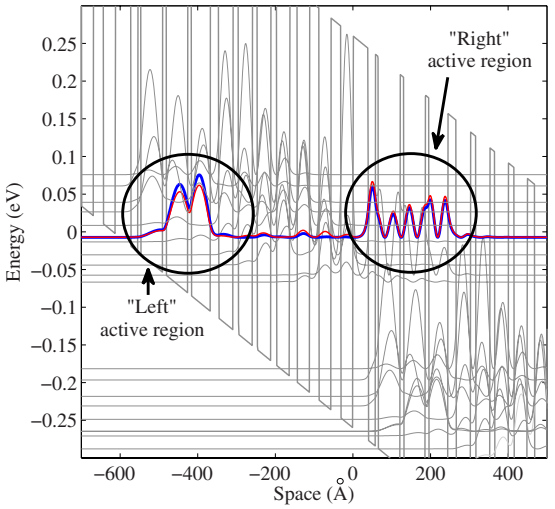


FIG. 2. (Color online) Conduction-band diagram and subband wave functions at a bias field of 42.14 kV/cm, where the current peak denoted by a star in Fig. 1 occurs. The two anticrossed wave functions responsible for the current peak are emphasized.

active region to the "right" one. Therefore once an electron falls into the lower state of the left active region, it immediately shows up in the upper state of the right active region. This shortcut results in a current spike, and is of course an artifact. Such artifacts are typical to models that exclude coherences.[4,19]

The results are completely changed when coherences are included. Figure 3 shows that as the two wave functions from Fig. 2 approach anticrossing, the coherence between them starts to build up. The coherence causes a destructive interference of the two anticrossed wave functions at the right active region and constructive interference at the left one. This interference completely cancels the current spike, as one can see in the thin continuous curve of Fig. 1.

The thin continuous curve in Fig. 1 shows however that even with coherences, there are still many current peaks, which are absent in the experiment.[18] This stems from the fact that the width of the current peak at anticrossing is inversely proportional to the dephasing time.[7] The only dephasing terms included so far are those obtained from Eq.

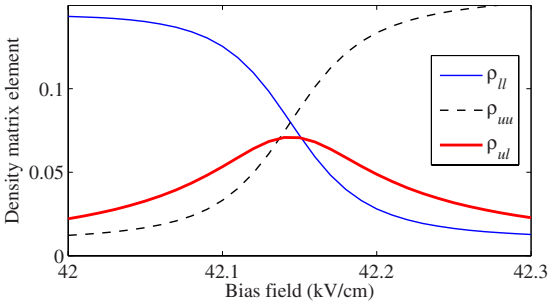


FIG. 3. (Color online) Density-matrix elements describing the two anticrossed states in Fig. 2. The subscripts $l$ and $u$ represent, respectively, the energetically lower- and upper-lying states. The coherence $\rho_{ul}$ reaches its maximum at the point where the wave functions $u$ and $l$ become most extended. $\mathrm{Im}(\rho_{ul})$, which is not shown, is two orders of magnitude smaller than $\mathrm{Re}(\rho_{ul})$.

(19) (e.g., terms of the form $R_{ijij}$), which originate from LO phonons and thus have a picosecond time scale. In reality there are faster dephasing mechanisms, such as the electron-electron ($e$-$e$) interaction.[10]

The method proposed in this work paves the way to include the $e$-$e$ dephasing mechanisms: one should follow the same steps with the $e$-$e$ scattering as we did with the phonon scattering. The dephasing terms originating from $e$-$e$ scattering will then be obtained automatically. Note that while the $e$-$e$ intersubband scattering is in general slow in QCLs,[20] it can become very fast when the energy separation between the subbands becomes small, which is exactly the situation at anticrossing.

Including the $e$-$e$ interaction in the model is beyond the scope of this paper. Here we just give a quick and simple phenomenological way to add dephasing. It is tempting to simply add the familiar second term on the right-hand side of Eq. (21) below to $R_{ijfg}$,

$$R'_{ijfg} = R_{ijfg} - \frac{\delta_{if}\delta_{jg}}{\tau} + \frac{\delta_{ig}\delta_{jf}}{\tau}, \tag{21}$$

where $\tau$ in Eq. (21) is a phenomenological dephasing time. However it is necessary to complement it by the last term, otherwise the basis invariance symmetry is broken and the equations lead to meaningless results.[21]

The results of Eq. (21), with $\tau = 50$ fs, are shown in Fig. 1 (thick blue curve). One can see that the added dephasing term reduces the current and smoothes off the peaks in it. This is what we expect from the Kazarinov-Suris theory. The blue curve in Fig. 1 is already qualitatively very similar to the experimental one.[18] Figure 1 clearly demonstrates that key role that coherences play in the transport of current across a QCL. Neglecting coherences leads to results that are clearly unphysical.

## V. DISCUSSION

To conclude, we have developed a theoretical approach to transport in QCLs. The existing incoherent QCL rate equations were extended to include coherence by imposing invariance of measurable quantities under basis change and positivity of the density matrix. The required computational effort is very modest. The key importance of coherence has been demonstrated. The discussion was dedicated to QCLs but the technique is applicable to heterostructures in general.

## APPENDIX A: BASIS INVARIANCE—FULL DERIVATION

We can now take $R_{iifg}$ as it is given in Eq. (13) and apply a similar basis transformation on its first two indices. Assuming that $i$, $j$, $f$, and $g$ are all different, Eq. (3) gives the relations

$$R_{ijfg} + R_{jifg} = R_{++fg} - R_{--fg}$$

$$R_{ijfg} - R_{jifg} = i[R_{+'+'fg} - R_{-'-'fg}], \tag{A1}$$

where the subscripts $+$, $+'$, $-$, and $-'$ refer now to

$$|+\rangle \equiv \frac{|i\rangle + |j\rangle}{\sqrt{2}}, \quad |-\rangle \equiv \frac{|i\rangle - |j\rangle}{\sqrt{2}}$$

$$|+'\rangle \equiv \frac{|i\rangle + i|j\rangle}{\sqrt{2}}, \quad |-'\rangle \equiv \frac{|i\rangle - i|j\rangle}{\sqrt{2}}. \tag{A2}$$

Adding the two Eqs. (A1) and substituting Eq. (13) at the right-hand side, we obtain

$$R_{ijfg} = Q^0_{ijfg}, \tag{A3}$$

where $Q^0_{ijfg}$ is given in Eq. (15).

Equation (A3) applies only when all four indices of $R$ are different from each other. In order to derive the $R$-$s$ with part of the indices equal, start with $R_{ffff}$, and apply on it the transformations according to Eqs. (9) and (11). After a tedious but straightforward calculation we obtain

$$R_{fgfg} + R_{gfgf} = -\Gamma_{+-} - \Gamma_{-+} - \Gamma_{+'-'} - \Gamma_{-'+'} - \frac{1}{2}\sum_{m\neq f,g} \Gamma_{+m} + \Gamma_{-m} + \Gamma_{+'m} + \Gamma_{-'m}. \tag{A4}$$

Equation (A4) gives the real part of the coefficient $R_{fgfg}$, which is commonly interpreted as dephasing. The imaginary part is interpreted as the change in the energy separation between levels $f$ and $g$ due to the interaction with the bath of phonons. Apparently our approach does not allow extracting this energy shift. The effect of phonons on subband energies at equilibrium is already included in $H$ since the commonly published material parameters already take them into account.[22] However our model apparently cannot predict nonequilibrium corrections to the energies. The imaginary part of $R_{fgfg}$ is henceforth neglected. Substituting Eqs. (4)–(7) into Eq. (A4), we thus obtain

$$R_{fgfg} = Q^0_{fgfg} - \frac{1}{2}\sum_m [Q^0_{ffmm} + Q^0_{ggmm}]. \tag{A5}$$

With Eq. (A5) at hand, we can derive coefficients of the form $R_{ijfj}$ as well. To this aim we employ Eq. (A3) again, where this time

$$|+\rangle \equiv \frac{|i\rangle + |f\rangle}{\sqrt{2}}, \quad |-\rangle \equiv \frac{|i\rangle - |f\rangle}{\sqrt{2}}$$

$$|+'\rangle \equiv \frac{|i\rangle + i|f\rangle}{\sqrt{2}}, \quad |-'\rangle \equiv \frac{|i\rangle - i|f\rangle}{\sqrt{2}}. \tag{A6}$$

We obtain

$$2R_{+j+j} = R_{ijij} + R_{fjfj} + R_{ijfj} + R_{fjij}.$$

$$2iR_{+'j+'j} = iR_{ijij} + iR_{fjfj} + R_{ijfj} - R_{fjij}. \tag{A7}$$

Adding up the two Eqs. (A7) one can express $R_{ijfj}$ in terms of $R_{ijij}$, $R_{fjfj}$, $R_{+j+j}$, and $R_{+'j+'j}$. All the latter coefficients can be calculated using Eq. (A5). The result is

$$R_{ijfj} = Q^0_{ijfj} - \frac{1}{2}\sum_m Q^0_{ifmm}. \tag{A8}$$

Collecting all the expressions derived here together, one arrives at Eq. (14).

## APPENDIX B: LINDBLAD FORM

Equation (1) preserves the trace and the non-negativity of the density matrix if and only if $R_{ijfg}$ can be written in the Lindblad form,[15,16]

$$R_{ijfg} = \sum_{a,b=1}^{N^2-1} c_{ab} R^{ab}_{ijfg}, \tag{B1}$$

where

$$R^{ab}_{ijfg} = F^a_{fi}F^{b*}_{gj} - \frac{\delta_{gj}}{2}\sum_{m=1}^{N} F^{b*}_{mf}F^a_{mi} - \frac{\delta_{if}}{2}\sum_{m=1}^{N} F^{b*}_{mj}F^a_{mg}. \tag{B2}$$

The theorem[15,16] strictly applies only to finite-dimensional Hilbert spaces. We thus imagine a space-discretized version of a QCL Hamiltonian, which is $N$ dimensional ($N$ being the number of spatial points). At the end we may let $N$ approach infinity so that the discretrized Hamiltonian approximates the continuous one arbitrarily well. In Eq. (B1) $c_{ab}$ is a Hermitian ($c_{ab}=c^*_{ba}$) positive-definite matrix of dimension $N^2-1$. $F^a$, with $1\leq a\leq N^2-1$, is a set of orthonormal traceless matrices

$$\mathrm{Tr}[F^a] = \sum_{i=1}^{N} F^a_{ii} = 0,$$

$$\mathrm{Tr}[F^{a\dagger}F^b] = \sum_{i,f=1}^{N} F^{*a}_{if}F^b_{if} = \delta_{ab}. \tag{B3}$$

By comparing Eq. (B3) with Eq. (19) one can see that the two equations coincide if $Q_{ijfg}$ can be written in the form

$$Q_{ijfg} = \sum_{a,b=1}^{N^2-1} c_{ab}F^a_{fi}F^{b*}_{gj}. \tag{B4}$$

In fact, we only need to show that the restriction of $Q_{ijfg}$ to the $N^2-1$-dimensional space of traceless matrices satisfies Eq. (B4) because the other part of $Q$ is anyway canceled out by Eq. (19). In order to show that such a restriction of $Q$ can be written as Eq. (B4), it is enough to show that $Q$ itself can be written as

$$Q_{ijfg} = \sum_{a,b=1}^{N^2} c_{ab}F^a_{fi}F^{b*}_{gj}, \tag{B5}$$

where this time we include $F^N$, which is proportional to the identity matrix. $F^N$ is the last basis matrix of the space of $N\times N$ matrices. If we show that $c_{ab}$ is positive definite in the $N^2$-dimensional space of $N\times N$ matrices, then its restriction to the $N^2-1$-dimensional space of traceless matrices is also positive definite.

Since the $F^a$ matrices are an orthonormal basis of the space of $N\times N$ matrices, any linear operator in this space can be written as the right-hand side of Eq. (B5). What we only need to show is that $c_{ab}=c_{ba}^*$, and that $c_{ab}$ is positive definite. The first follows from the property $Q_{ijfg}=Q_{jigf}^*$, which is easily seen from Eqs. (5) and (20). To show that $c_{ab}$ is positive definite, we have to show that given an arbitrary matrix $w_{if}$,

$$\sum_{i,j,f,g=1}^{N} w_{jg}^* Q_{ijfg} w_{if} = \sum_{ijfg,k} A_{if}(k) w_{if} A_{jg}^*(k) w_{jg}^* \sqrt{g_{if}(k) g_{jg}(k)} = \sum_k \left| \sum_{if} A_{if}(k) w_{if} \sqrt{g_{if}(k)} \right|^2 \quad \text{(B6)}$$

is positive. In the second line of Eq. (B6) we used Eq. (20). The integration over $k$ has been replaced by a sum because we have switched to a discrete Hamiltonian. From the last line of Eq. (B6) it is evident that indeed $c_{ab}$ is positive definite.

In order to examine how much we are allowed to deviate from Eqs. (19) and (20), we note that $Q_{ijfg}$ must satisfy the inequality

$$Q_{ijfg}^2 \le Q_{iiff} Q_{jjgg}. \quad \text{(B7)}$$

Equation (B7) follows from Eq. (B4) [or Eq. (B5)] and the Cauchy-Schwartz inequality since $Q_{ijfg}$ can be viewed as a scalar product between the two vectors $F_{fi}^a$ and $F_{gj}^a$. This is allowed because $c_{ab}$ is positive definite and can thus be viewed as a metric tensor.

Let us now assume that we accept the wave-function-dependent part of Eq. (20), which was argued for at the end Sec III, and allow a general energy-dependent part

$$Q_{ijfg} = \sum_k A_{if}(k) A_{jg}^*(k) G_{ijfg}(k). \quad \text{(B8)}$$

Multiplying and dividing Eq. (B8) by $\sqrt{g_{if}(k)}$ and invoking the Cauchy-Schwartz inequality, we obtain

$$Q_{ijfg}^2 \le Q_{iiff} \sum_k |A_{jg}^*(k)|^2 \frac{G_{ijfg}^2(k)}{g_{if}(k)}. \quad \text{(B9)}$$

A sufficient condition for Eq. (B7) to hold is therefore that

$$G_{ijfg}^2(k) \le g_{if}(k) g_{jg}(k). \quad \text{(B10)}$$

We argue that Eq. (B10) is also a necessary condition, because if Eq. (B10) is violated at some $k_0$, we can imagine a set of wave functions where $A_{if}(k)$ and $A_{jg}(k)$ are sharply peaked around $k_0$. Then Eq. (B7) is violated. Since we assumed that $G_{ijfg}^2(k)$ depends only on the energies but not on the wave functions, we are allowed require Eq. (B7) to hold for any choice of wave functions.

Since Eq. (B10) reaches equality when $\varepsilon_i=\varepsilon_j$ and $\varepsilon_f=\varepsilon_g$ [see Eq. (15)], the lowest-order correction to Eq. (15) in $(\varepsilon_i-\varepsilon_j)$ and $(\varepsilon_f-\varepsilon_g)$ must be at least of second order

$$\frac{G_{ijfg}(k)}{\sqrt{g_{if}(k) g_{jg}(k)}} = 1 - \alpha(k)[(\varepsilon_i-\varepsilon_j)^2 + (\varepsilon_f-\varepsilon_g)^2] \quad \text{(B11)}$$

plus higher-order terms, where $\alpha(k)$ is a positive function of $k$. First-order corrections, or a mixed $(\varepsilon_i-\varepsilon_j)(\varepsilon_f-\varepsilon_g)$ term, will violate Eq. (B10). Since coherences are nonvanishing only in the vicinity of degeneracy, this second-order correction in the distance from degeneracy is likely to be negligible.

*gordon_ariel@yahoo.com


[1] J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, and A. Y. Cho, Science **264**, 553 (1994).

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

[3] P. Harrison, D. Indjin, I. Savic, Z. Ikonic, C. A. Evans, N. Vukmirovic, R. W. Kelsall, J. McTavish, V. D. Jovanovic, and V. Milanovic, Proc. SPIE **6909**, 690912 (2008).

[4] H. Callebaut and Q. Hu, J. Appl. Phys. **98**, 104505 (2005).

[5] A. Wacker, Phys. Status Solidi C **5**, 215 (2008).

[6] C. Weber, A. Wacker, and A. Knorr, Phys. Rev. B **79**, 165322 (2009).

[7] R. F. Kazarinov and R. A. Suris, Sov. Phys. Semicond. **5**, 207 (1971).

[8] D. Indjin, P. Harrison, R. W. Kelsall, and Z. Ikonić, J. Appl. Phys. **91**, 9019 (2002).

[9] J. C. Tolédano and P. Tolédano, *The Landau Theory of Phase Transitions* (World Scientific, Singapore, 1987).

[10] P. Harrison, D. Indjin, and R. W. Kelsall, J. Appl. Phys. **92**, 6921 (2002).

[11] M. Troccoli, G. Scamarcio, V. Spagnolo, A. Tredicucci, C. Gmachl, F. Capasso, D. L. Sivco, and M. Striccoli, Appl. Phys. Lett. **77**, 1088 (2000).

[12] D. Ahn and S. L. Chuang, Phys. Rev. B **37**, 2529 (1988).

[13] At least without in-plane magnetic fields.

[14] One can then extend the analysis to show that coherences cannot excite each other by examining two equations with two coupled coherences and slow source terms.

[15] G. Lindblad, Commun. Math. Phys. **48**, 119 (1976).

[16] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, J. Math. Phys. **17**, 821 (1976).

[17] R. W. Boyd, *Nonlinear Optics*, 2nd ed. (Academic, London, 2003).

[18] J. S. Yu, S. Slivken, A. Evans, S. R. Darvish, J. Nguyen, and M. Razeghi, Appl. Phys. Lett. **88**, 091113 (2006).

[19] H. Li, J. C. Cao, and J. T. Lü, J. Appl. Phys. **103**, 103113 (2008).

[20] P. Harrison, Appl. Phys. Lett. **75**, 2800 (1999).

[21] Actually even Eq. (21) is only invariant under *real* basis transformations and is not entirely compatible with the Kazarinov-Suris theory. This is enough for a qualitative illustration but not for quantitative calculations.

[22] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, J. Appl. Phys. **89**, 5815 (2001).