Self-consistent energy balance simulations of hole dynamics in SiGe/Si THz quantum cascade structures

Z. Ikonić,[a] P. Harrison, and R. W. Kelsall
Institute of Microwaves and Photonics, School of Electronic and Electrical Engineering, University of Leeds, Leeds LS2 9JT, United Kingdom

(Received 21 July 2004; accepted 17 August 2004)

Analysis of hole transport in cascaded *p*-Si/SiGe quantum well structures is performed using self-consistent rate equations simulations. The hole subband structure is calculated using the 6×6 $\mathbf{k}\cdot\mathbf{p}$ model, and then used to find carrier relaxation rates due to the alloy disorder, acoustic, and optical phonon scattering, as well as hole-hole scattering. The simulation accounts for the in-plane $\mathbf{k}$-space anisotropy of both the hole subband structure and the scattering rates. Results are presented for prototype THz Si/SiGe quantum cascade structures. © *2004 American Institute of Physics*. [DOI: 10.1063/1.1805727]

I. INTRODUCTION

The development of III-V semiconductors based intersubband transition quantum cascade lasers (QCL) emitting in the mid- and, more recently, far-infrared (THz) range, has intensified research efforts to realize analogous devices in Si/SiGe strained-layer technology. The possibility of monolithic integration of silicon based electronic and optoelectronic components is a strong incentive for this research. It is presently considered that *p*-type Si/SiGe structures are more promising candidates for QCLs than *n*-type ones, because the valence band takes a larger share of the band gap discontinuity at heterostructure interfaces. Evolving from the earlier proposals of suitable structure,[1,2] the present status of the research includes the demonstration of successful growth of long Si/SiGe cascades, and observations of electroluminescence in both the mid-infrared[3,4] and far-infrared[5,6] wavelength ranges. Bias-tunable emission wavelength has also been obtained,[7] although full laser operation has yet to be achieved.

The gain depends on the scattering rates between different subbands and also between different in-plane momentum states within a subband (carrier heating/cooling effects). These effects have been extensively studied in QCLs based on conduction band intersubband transitions. One approach relies on self-consistent solutions of rate equations.[8–13] Another approach uses the microscopic, and computationally more demanding Monte Carlo technique.[14–16] Although the latter does not make the assumption of equilibriumlike carrier distributions over states within any single subband, and therefore gives a deeper insight into the electron dynamics, the former are much faster while still giving quite good estimates of device characteristics.

While there have been previous theoretical studies of hole transport in quantum confined Si/SiGe systems,[17,18] based on a fully anisotropic 6×6 $\mathbf{k}\cdot\mathbf{p}$ description of the subband structure, these have focused on in-plane transport as is relevant for microelectronic devices, rather than on vertical transport as is of interest for quantum cascade structures. We have recently reported on Monte Carlo simulations of hole transport in *p*-Si/SiGe cascades, but with hole-hole scattering neglected.[19] The problem is generally similar to the case of *n*-type cascades, but is more complex because the subband structure and scattering rates are anisotropic and strongly dependent on the in-plane momentum of the hole states. Here we describe the development and implementation of a self-consistent energy balance method for calculating hole dynamics in *p*-Si/SiGe quantum cascade structures, with hole-hole scattering included.

II. CALCULATION DETAILS

A. Hole subband structure and scattering rate calculation

The calculation of hole subband structure and single-hole scattering rates have been described in detail previously,[19] and are therefore summarized only briefly here, and more extensive details are provided for the calculation of hole-hole scattering, and for the self-consistent energy-balance approach. The subbands of a biased cascade are found by solving the Schrödinger equation in a structure with a limited number of unit cells, or periods (each of which may be structurally simple, comprising a single quantum well and barrier, or quite complex), subject to either box or periodic boundary conditions. Their wave functions are usually localized in a single period of the cascade. In a long cascade the states show quasiperiodicity; i.e., by translating the wave function of a state by one period, and shifting its energy by the potential drop across one period, another actual state of the system (the one mostly localized in the next period) is obtained. Having a set of states assigned to a particular (hereafter called the central) period, the corresponding sets assigned to other periods may be constructed simply by using quasiperiodicity.

The hole band structure is calculated using the 6×6 $\mathbf{k}\cdot\mathbf{p}$ scheme in the plane wave implementation, as described previously,[20] using Foreman's boundary conditions.[21] The accuracy of the method is good for the class of structures of interest in this work.[22] In some cases, such as mid-infrared cascades which involve higher transition energies

[a]Electronic mail: eenzi@leeds.ac.uk

and therefore larger in-plane wave vectors ($\mathbf{k}_{\|}$), it may be necessary to include more bands and use the recently developed 14×14 $\mathbf{k}\cdot\mathbf{p}$ model.[23] For the scattering rate calculations one would then also need the hole-phonon coupling Hamiltonian and its parameters, beyond those established for the 6×6 $\mathbf{k}\cdot\mathbf{p}$ model, so the latter seems a reasonable choice at present. Indeed, much of the recent work on hole mobility in Si or SiGe, of either Monte Carlo or scattering-rate type, was also carried out within the 6×6 model.[17,18] Mixing of heavy-hole (*hh*), light-hole (*lh*), and spin-orbit split-off (*so*) valence band states results in both a prominent in-plane nonparabolicity and anisotropy of hole subbands, so the energies and wave functions of the subbands of interest are tabulated at a number of $\mathbf{k}_{\|}$ values. This is done in the irreducible wedge of the two-dimensional (2D) Brillouin zone (for the usual, [001] grown structures this is 1/8 of the full 2D Brillouin zone), and the symmetry properties are used to reconstruct states outside this wedge by rotation (in contrast, for parabolic-dispersion electronic subbands it suffices to find states at $\mathbf{k}_{\|}=0$ only). Tracking particle dynamics in a cascade requires all the states to be assigned to individual periods, generally based on the wave function localization. Special care is taken to avoid double counting when hybridization i.e., anticrossing of states otherwise localized in two different periods, appears[19] (for holes this is a $\mathbf{k}_{\|}$-dependent phenomenon, because states which are well separated in energy, e.g., at $\mathbf{k}_{\|}=0$ may come closer and anticross at some other value of $\mathbf{k}_{\|}$, which swaps their localization properties).

The main inelastic scattering mechanism for holes in SiGe is the deformation potential scattering (via acoustic and nonpolar optical modes). Optical phonon scattering in the alloy layers is described by assuming three distinct modes, corresponding to Ge-Ge, Ge-Si, and Si-Si interatomic vibrations, each with its own frequency and deformation potential, as well as an appropriate weight, according to the number of interatomic bonds present in the alloy.[24] For acoustic phonons, on the other hand, the weighted averages of the sound velocities and deformation potentials of Si and Ge are taken. The phonons are considered to be bulklike, and the tensorial, rather than scalar, form of the hole-phonon interaction Hamiltonian is used.[25–29] Optical phonons are assumed to be nondispersive, while acoustic phonons are taken to have linear dispersion; i.e., the quasielastic approximation is not used. Another important scattering mechanism in SiGe is the alloy disorder scattering, which is purely elastic, but nevertheless can induce hole transitions between different subbands.[30–32,20] Due to the nonparabolicity and anisotropy of the hole subbands, the scattering rates are evaluated numerically, using the linear tetrahedron method[17,33–36] of appropriate dimensionality.

Upon calculating the microscopic (mesh-cell to mesh-cell) scattering rates, these are averaged[37] by weight factors corresponding to the population of the intial state, and accounting for the population of the final state (via the Pauli exclusion principle). Fermi-Dirac distribution functions (with independent carrier temperatures, generally different from the lattice temperature) are used for this purpose. The averaged scattering rates then depend on carrier densities in both the initial and final states. However, direct calculations show that this dependence becomes significant only at very low values of carrier (not lattice) temperature, or large (degenerate) densities of carriers, none of which is encountered in cascade structures of practical interest. This may not always be true for electrons, where the effective mass is small and the Fermi level may be close to, or above, the subband minima. Holes have a considerably larger effective mass and the Fermi level is, under usual conditions in cascades, sufficiently below the subband minima to imply relatively small occupancy of states. Therefore, the averaged scattering rates may be taken, to a very good approximation, to be independent of actual carrier densities, as well as of the final state temperature, while depending on the temperature of the initial state.

Carrier-carrier scattering (Auger interaction) has been recognized as an important mechanism in quantum cascade lasers, particularly in devices with closely spaced subbands. This is a two-body process, in which holes in the initial states $(i,\mathbf{k}_{\|i})$ and $(j,\mathbf{k}_{\|j})$ scatter into final states $(f,\mathbf{k}_{\|f})$ and $(g,\mathbf{k}_{\|g})$, where some or all of the state indices may be the same. Electron-electron scattering in QCLs has been extensively studied,[38,39] assuming single subband static screening within the random phase approximation (RPA). Simulations of actual cascade structures which include electron-electron scattering calculated in this manner show quite good agreement with experiments.[10–12] Although the validity of this screening model may seem questionable in view of the results of calculations which use the dynamic RPA dielectric function,[40] the latter is computationally too demanding to be manageable for structures of the complexity encountered in present QCLs. For holes, with their nonparabolicity and anisotropy, the dynamic screening approach would be even slower. In this work we have therefore used static screening, the only modifications being the replacement of the conventional, truly single subband, by "weight-averaged" static screening,[41] and accounting for the spinor structure of hole wave functions when evaluating the interaction matrix elements. The hole-hole scattering rate is thus calculated from Eq. (49) in Ref. 38; or Eq. (2.1) in Ref. 39. Previous calculations of intersubband Auger processes for holes in SiGe single quantum wells (for photodetectors)[42] (using a simple, somewhat arbitrarily chosen, constant screening factor) indicate that this may be a fast process. It is interesting to note that, due to the mixed parity of hole subbands for general $\mathbf{k}_{\|}$ even in symmetric structures, there are no selection rules for the subband indices for hole-hole scattering, in contrast to the case of electrons.[39] The expression for carrier-carrier scattering is usually written in a somewhat asymmetric manner, giving the scattering rate of a particle in state $(i,\mathbf{k}_{\|i})$ with all particles in subband j, but any value of $\mathbf{k}_{\|j}$. Evaluation would then require the 4D linear tetrahedron method.[36,43] However, to find microscopic scattering rates we fix the wave vectors $\mathbf{k}_{\|i}$ and $\mathbf{k}_{\|j}$ of both initial states (to the centers of their mesh cells), and use the 2D tetrahedron method.

Concerning the scaling properties of hole-hole scattering, as the hole density varies, conclusions similar to those for single-hole scattering processes apply. Except at very low temperatures and very high densities, the total scattering rate from the pair of initial states i and j into final states f and g

is proportional, to a very good approximation, to the product of populations of these states, $n_i n_j$, and the proportionality constant $w_{i,j,f,g}$ then depends on hole temperature(s) in these two subbands but not on any of the populations.

B. Rate equations in a cascade

Carrier transport in a quantum cascade structure is described within a tight-binding-like picture. Any state in a long cascade is associated to one of its periods. A biased periodic cascade, assuming a globally linear variation of the internal electrostatic potential (i.e., no domain formation), has a self-similar potential, invariant upon translation by a period D and energy shift by the potential drop across a period ΔV, Consequently, if $\psi(z)$ is a solution of the Schrödinger equation at energy E, then $\psi(z-D)$ is a solution at $E-\Delta V$ (quasiperiodicity). This allows all the states in a cascade to be constructed as replicas, shifted in space and energy, of the set of states assigned to one of its periods. Among the states actually calculated in a structure with a finite number of periods, those which are mostly localized near the middle of the structure are clearly most trustworthy as representatives of states in an infinite cascade, because they are sufficiently remote from the boundaries, and are used in the "replication" process.[19] To keep the problem tractable the number of states N assigned to a single period is limited, based on the expectation that high-energy states will be virtually unpopulated, because most of the carriers will scatter into lower energy states of subsequent periods rather than remain in ever higher states as they move along the cascade. States assigned to the central period will be labeled as ψ_i, $(i=1,\ldots,N)$, and those assigned (by construction) to its kth nearest neighbor will be labeled as ψ_{i+kN}, $(i=1,\ldots,N)$, where $k>0$ for right and $k<0$ for left neighbors. The number of neighboring periods taken in any practical calculation is also limited to some value (M), based on the fact that there is normally a very small overlap of wave functions belonging to more distant periods, resulting in negligible scattering between such states.

Having calculated all the wave functions and scattering rates, the system of rate equations can be written as

$$\frac{dn_f}{dt} = \sum_i n_i w_{i,f} - n_f \sum_i w_{f,i} + \sum_{i,j,g} n_i n_j w_{i,j,f,g} - n_f \sum_{i,j,g} n_g w_{f,g,i,j}, \tag{1}$$

where i, j, f, and g run over all states in the cascade, in all of its periods.

The assumption of a globally linear potential variation (no bowing) inside the cascade enables the use of "periodic boundary conditions" for the particle densities: these are taken to be identical in all periods of the cascade, i.e., $n_i = n_{i+kN}$ (where $i=1,\ldots,N$ and $k=\pm1,\pm2,\ldots$), and the total density equals the sheet doping density per period n_{tot}. It then suffices to account explicitly for particles in a single period, with a relatively small number of states, rather than in the whole cascade. The rate equations may then be written as

$$\frac{dn_f}{dt} = \sum_{k=-M}^{M}\sum_{i=1}^{N} n_i w_{i+kN,f} - n_f \sum_{k=-M}^{N}\sum_{i=1}^{N} w_{f,i+kN} + \sum_{\substack{k,k',k''=-M \\ (|\Delta k|\le M)}}^{M} \left[\sum_{i,j,g=1}^{N} n_i n_j w_{i+kN,j+k'N,f,g+k''N} - n_f \sum_{i,j,g=1}^{N} n_g w_{f,g+k''N,i+kN,jk'N} \right]. \tag{2}$$

Furthermore, the scattering rates are shift invariant: e.g., for single particle scattering rates we have $w_{i,j}=w_{i+kN,j+kN}$, or $w_{i-N,j}=w_{i,j+N}$, and similarly for the hole-hole scattering coefficients. Although the scattering-rate subscripts in Eq. (2) take both positive and negative values, it is always possible to use the shift-invariance property and make all the subscripts positive. For instance, the first two terms in Eq. (2), which describe single-hole scattering, may then be written as

$$\sum_{i=1}^{N} n_i w_{i,f} - n_f \sum_{i=1}^{N} w_{f,i} + \sum_{k=1}^{M} \left[\sum_{i=1}^{N} n_i (w_{i,f+kN} + w_{i+kN,f}) - n_f \sum_{i=1}^{N} (w_{f,i+kN} + w_{f+kN,i}) \right]. \tag{3}$$

Therefore, only the scattering rates within the central period, and those linking it to its right neighbors (up to the desired order) need be evaluated. Wave functions of the left neighbors are not explicitly required. In calculating the central cell wave functions, however, it is still advisable to have all the left neighboring periods present, for the purpose of finding these wave functions more accurately, and in their full spatial extent as necessary for evaluation of "overlap" type integrals for scattering. This is a sufficient condition to warrant accuracy under all circumstances (e.g., the occurence of hybridization between some of central-period states and states of remote periods). In conduction-band cascades it is simple to check, by visual inspection of wave functions, whether hybridization of remote states occurs or not, and in the latter case it may be sufficient to consider just 2 or $2\frac{1}{2}$ periods. However, in valence-band cascades hybridization may appear for some $\mathbf{k}_{\|}\neq 0$ even if it was not observed at $\mathbf{k}_{\|}=0$, and the wave function computational window should thus cover both the left and right neighboring periods.

Out of the total of N, there are $N-1$ linearly independent equations, so one of them is replaced by the particle conservation law: $\Sigma n_i = n_{\mathrm{tot}}$. If the carrier temperatures, and hence the scattering rates, are known, Eqs. (2) may be solved in the steady state $(d/dt=0)$ or, given the initial conditions, in the nonstationary case. Although the system is nonlinear if hole-hole scattering is included, solving is still very fast compared to the cost of evaluating all the scattering rates, due to the assumption that the rates $w_{i,j}$ and $w_{i,j,f,g}$ do not depend on the carrier densities, which results in a very simple analytic model dependence on the unknowns. This approach constitutes the particle rate equation model, which needs the carrier temperatures as input, either via an "educated guess" or simply set equal to the lattice temperature.

A more refined model accounts for the fact that subbands exchange energy, as well as particles, in all the scattering processes—elastic or inelastic. The kinetic energy transfer rates between subbands can be found in a similar manner as the particle scattering rates, where the kinetic energy is defined as the hole energy at wave vector $\mathbf{k}_{\|}$ relative to the minimal value in the same subband (usually at the zone center). A hole leaving subband i takes its own energy out of the ensemble in this subband, while the ensemble in the subband f in which the hole arrives increases its total energy by the kinetic energy corresponding to the hole wave vector in this subband. The quantized parts of the subband energies, and the energy of the scatterer (nonzero in case of phonons) here have the role of potential energy. There is no conservation of either the kinetic or the full energy of the hole subsystem, because energy may be exchanged with the lattice. Here the two kinetic energies are completely unrelated to each other, and depend on the scattering process which has taken place. Therefore, instead of a single energy transfer rate that occurs in the $i \rightarrow f$ transitions, two rates are defined: w_{if}^{e+} and w_{if}^{e-}, which denote the kinetic energy increment rate in subband f, and the kinetic if energy decrement rate in subband i, respectively, due to $i \rightarrow f$ transitions. Within the nondegenerate approximation both rates depend only on the hole temperature in subband i and the total rates are proportional to the hole density in subband i. The kinetic energy rate equations for the total kinetic energy E_f^{tot} stored in a state f read

$$\frac{dE_f^{\mathrm{tot}}}{dt} = \sum_i n_i w_{if}^{e+} - n_f \sum_i w_{fi}^{e-} + \sum_{i,j,g} n_i n_j w_{ijfg}^{e+} - n_f \sum_{i,j,g} w_{fgij}^{e-} n_g, \tag{4}$$

and can then be written in terms of neighbor-order contributions, in the same manner as in Eq. (2). All N equations are linearly independent.

Similar considerations for conduction subbands in cascades have been presented previously,[13] but the electron temperature was assumed equal in all subbands, while the hole subband temperatures here are considered as independent variables. The case of a single carrier temperature, common for all subbands but distinct from the lattice temperature, is formally recovered by summing Eq. (4) over all states of a period, and using the single carrier temperature in evaluating all $w_{i,j}$ and $w_{i,j,f,g}$ terms.

It should be noted that intrasubband scattering rates, where particle(s) remain in the initial subband, do not enter (i.e., are cancelled out) in the particle-density rate equations (2). However, they are present in the kinetic energy rate equations (4), because carriers may exchange energy with the lattice via inelastic scattering mechanisms, and the net flow of energy is nonzero if the carrier temperature differs from that of the lattice. There is another role of intrasubband scattering processes in particle-density rate equations, which is only implicit at the rate equation level of description. This is to drive the particle distributions towards their equilibriumlike (Fermi-Dirac) forms, on which the rate equation approach relies. The fastest mechanism which performs this function is the intrasubband carrier-carrier scattering[44] (with a temperature unrelated to the lattice temperature, however), while the generally slower carrier-phonon scattering drives the carrier temperature towards the lattice value.

In quantum cascade structures only the scattering between adjacent periods is normally considered ("nearest neighbor approximation"[15]), and only when including higher, more continuumlike, states, or in photodetectors, may it become necessary to include interactions with more distant states. If scattering to all neighbors is taken into account, the precise way of state assignment to individual periods would in fact be irrelevant, because all the scattering processes would be accounted for anyway. However, if interaction with only a limited number of nearest neighbors is to be taken, proper assignment becomes very important, because this determines which states are remote in real space, such that scattering between them can be justifiably neglected. In this work we consider only the nearest neighbor interactions. For hole-hole scattering, in accordance with Eq. (2), we interpret this to mean that the initial and final states of *both* carriers belong to at most two (not three) adjacent periods.

In the steady state the system (2) and (4), which constitutes the self-consistent energy-balance (SCEB) model, is a set of $2N$ nonlinear equations in N values of densities and N temperatures. In these equations the density dependence is extracted in analytic form, because this is a very good approximation to the true dependence which would be calculated numerically, while the dependence on temperatures (which control the scattering rates) must be evaluated numerically. Solving this system is therefore a difficult numerical problem, and is performed iteratively, using the Broyden method. It does not require the derivative information at start, but rather builds successively better approximations to the Jacobian as it walks through the space of variables. However, as is quite common in solving nonlinear sets of equations, the outcome may not always be the true solution; i.e., the code may settle in a local minimum, which depends on the initial guess (starting point of the procedure). If the initial guess is very remote from the solution we find that it is unlikely the solution will be reached in any reasonable number of iterations, if at all, and the procedure has to be restarted. In such situations it is helpful to find first the single carrier-temperature solution of the system, which can be obtained using simple bisection for the temperature (which always converges) in conjunction with the fast solution of the particle rate equations at each temperature point. Using this solution as the input to Broyden procedure for the full SCEB calculation greatly improves the chances of finding the complete solution.

By observing that particle conservation in a period may be written as $\Sigma_{i=1}^{N}(\sqrt{n_i})^2 = n_{\mathrm{tot}}$, it is clear that N densities can be expressed in terms of $N-1$ angles ϕ_i, $i=1,\ldots,N-1$, by hyperspherical parametrization

$$\sqrt{n_1} = R\cos(\phi_1)$$

$$\sqrt{n_2} = R\sin(\phi_1)\cos(\phi_2)$$

$$\sqrt{n_3} = R\sin(\phi_1)\sin(\phi_2)\cos(\phi_3)$$

$$\vdots$$

$$\sqrt{n_{N-1}} = R\,\sin(\phi_1)\ldots\sin(\phi_{N-2})\cos(\phi_{N-1})$$

$$\sqrt{n_N} = R\,\sin(\phi_1)\ldots\sin(\phi_{N-2})\sin(\phi_{N-1}), \tag{5}$$

where $R=\sqrt{n_{\mathrm{tot}}}$. If all the densities are to be non-negative, the angles have to satisfy $0\leqslant\phi_i\leqslant\pi/2$, and any point within such a hypercube gives a different, physically allowed set of densities n_i. The order of the system is then reduced to $2N-1$. Within this implementation the Broyden procedure goes along a different path than within the direct implementation (of order $2N$). While both of them usually work well, we have encountered examples where one or the other failed to converge to the solution, and so care should be taken when using either approach.

The current density is evaluated by accounting for all carriers which pass through some reference plane, e.g., the interface between the central period and the adjacent period to the right, and is given by

$$J = \sum_{\substack{k,k'=-M \\ (|\Delta k|\leqslant M)}}^{M} \sum_{i,f=1}^{N} n_i w_{i+kN,f+k'N}[\Theta(k')-\Theta(k)]$$
$$+ \sum_{\substack{k,k',k''k'''=-M \\ (|\Delta k|\leqslant M)}}^{M} \sum_{i,j,f,g=1}^{N} n_i n_i w_{i+kN,j+k'N,f+k''n,g+k'''N}$$
$$[\Theta(k'')+\Theta(k''')-\Theta(k)-\Theta(k')], \tag{6}$$

where $\Theta(k)=1$ if $k\in(1,M)$, and $\Theta(k)=0$ if $k\in(-M,0)$. Again, using the shift invariance of scattering rates helps to simplify this expression, e.g., the first term (due to single-hole scattering) may then be written as

$$\sum_{i=1}^{N} n_i \sum_{k=1}^{M} k \sum_{f=1}^{N} (w_{i,f+kN} - w_{i+kn,f}). \tag{7}$$

III. NUMERICAL RESULTS AND DISCUSSION

Self-consistent energy balance simulations have been performed for several *p*-Si/SiGe quantum cascade structures. In the band structure calculation the material parameters for Si and Ge were take from Refs. 45 and 46. The acoustic and optical phonon deformation potentials were taken from Ref. 27 (set C), and the alloy scattering potential was set to 0.3 eV (normalized to the primitive cell volume).[32,47]

The first example is a cascade of the simplest possible structure; i.e., a heterostructure stack of alternating wells and barriers. It has 16 monolayer (4.41 nm) $Ge_{0.3}Si_{0.7}$ wells and 8 monolayer (2.15 nm) wide Si barriers, grown on a $Ge_{0.2}Si_{0.8}$ virtual substrate. The structure is strain balanced, and can in principle be grown with an arbitrarily large number of periods. It has just two low-lying states per period, the ground HH1 and the first excited, LH1 state; the next, HH2 state is sufficiently higher in energy to remain almost inaccessible to holes throughout the range of biases used in the calculation. The LH1-HH1 energy spacing in the Si/SiGe system is primarily determined by the strain in the quantum well layers, and here amounts to 27.5 meV. In a biased cascade the alignment of the HH1 state from the preceding (higher) well and the LH1 state of the next (lower) well at $\mathbf{k}_\parallel=0$ occurs at a field of 42 kV/cm. However, for finite $\mathbf{k}_\parallel$ the alignment appears at different fields, because of the different dispersions of the HH and LH states, so the phenomenon of resonance is not so strong as in the case of *n*-type heterostructures. As the bias field varies, the spacing between LH1 and HH1 states of the same well changes only slightly, and most of the potential drop per period manifests in the displacement of the sets of states belonging to adjacent periods. The position of states at 60 kV/cm bias is shown in Fig. 1.

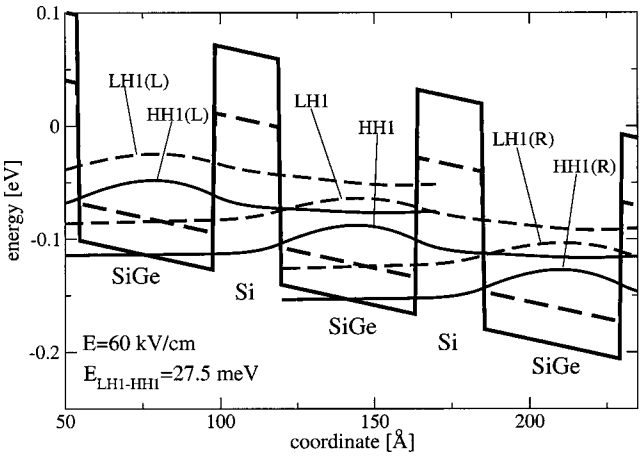

FIG. 1. Relevant hole states in the central and the two adjacent periods of a *p*-Si/SiGe cascade comprising 16 monolayer (4.41 nm) $Ge_{0.3}Si_{0.7}$ wells and 8 monolayer (2.15 nm) Si barriers, grown on a $Ge_{0.2}Si_{0.8}$ virtual substrate. The wave functions of HH and LH states, as well as the valence band edges for heavy and light holes (different in this strained system), are denoted by solid and dashed lines, respectively. The labels (L) and (R) denote states localized in wells lying to the left and right, respectively, of the central well.

Such a structure offers potential as an intersubband THz laser, where the lasing transition would be the interwell (diagonal) HH1$\rightarrow$LH1(R) transition, while the intrawell (vertical) LH1$\rightarrow$HH1 is the relaxation transition, emptying the lower laser state (see Fig. 1 for notation). There are two main reasons for using an interwell optical transition: (i) optical matrix elements for both in-plane and normally polarized light are larger for the interwell than for the intrawell HH1-LH1 transition, and (ii) it is plausible to expect that the upper state will be less populated than the lower state in the same well, which would imply that a population inversion would automatically exist for the interwell transition. The HH1$\rightarrow$LH1(R) interwell transition energy can be tuned by the bias field, which is an attractive feature: at $\mathbf{k}_\parallel=0$ it is zero at 42 kV/cm, and increases by 6.6 meV for each 10 kV/cm of excess bias.

The results of simulations of such a quantum cascade structure are shown in Figs. 2–7. The LH1 state population and the current density, calculated either within the multiple-temperature or single-temperature models, or with the carrier temperatures set equal to the lattice temperature, are given in Fig. 2. There may be a significant difference between the predictions of different models, particularly at lower values of the lattice temperature, which points to the importance of finding the carrier temperature via the fully self-consistent

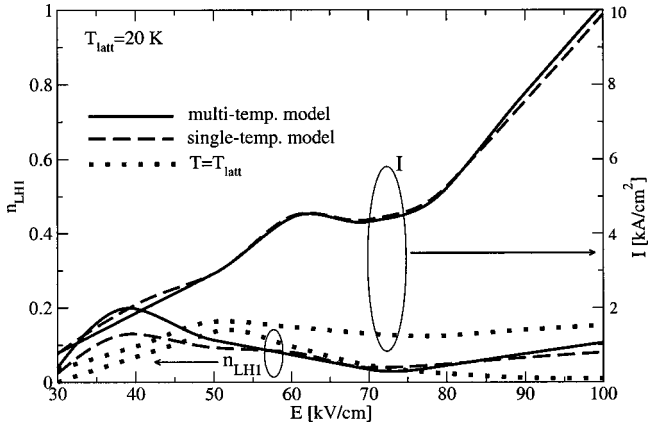

FIG. 2. The bias dependence of the LH1 subband population and the current density in the *p*-SiGe cascade described in Fig. 1, calculated without hole-hole scattering, at $T_{latt}=20$ K.

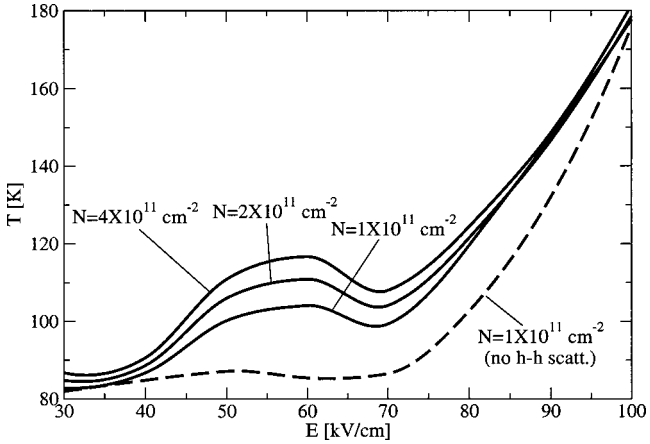

FIG. 4. The bias dependence of hole temperature in the *p*-SiGe cascade described in Fig. 1, calculated within the single-temperature model, for different values of hole sheet density per period, at $T_{latt}=20$ K.

energy balance solution. Both the LH1 state population and the current have peaks in the broad range of the HH1-LH1(*R*) alignment bias (hence the population inversion is worst there, but this is not the working bias for lasing anyway, as explained above). In Fig. 3 we plot the carrier temperatures, calculated both within the single-temperature and multiple-temperature models, without hole-hole scattering, for different values of the lattice temperature. Carrier temperatures calculated within the single-temperature model depend on the lattice temperature, though not in a simple additive manner; i.e., the carrier heating decreases with increasing lattice temperature. Similarly to n_{LH1} and *I*, the temperatures calculated within the multiple-temperature model, but not the single-temperature model, show characteristic features around the alignment bias. Although these originate from microscopic scattering rates as they are, it is also possible to give a more "macroscopic" explanation. Near the alignment bias the injection from HH1 into the LH1 of the next well is efficient, and the hole density in LH1 is large. The intrawell HH1-LH1 spacing in this structure is only a little less than the smallest (Ge-Ge) optical phonon energy, so that a relatively small increase in the LH1 subband temperature, relative to the lattice temperature, will start to open this channel of heat dissipation into the lattice—but this need not be excessive, because the electrical power input is still small. At larger biases, however, the carrier temperatures increase more or less monotonically: carriers have to heat up, and open all channels of heat dissipation, in order to accomodate the increased electrical power input. It should be noted that the individual temperatures of the two subbands may have very dissimilar values and behavior with varying bias, but their weighted average (in respect to the subband populations) roughly corresponds to the value obtained from the single carriers temperature model. It is also interesting to note that the agreement between the results of Monte Carlo[19] and self-consistent energy balance simulations (both with hole-hole scattering neglected) is reasonable in terms of the subband populations and current density, but the Monte Carlo method predicts considerably larger carrier temperatures. Similar differences between the carrier temperatures have been found in simulations of *n*-GaAs/AlGaAs cascades.[13,14]

The influence of hole-hole scattering, for different values of the total hole sheet density in this cascade is shown in

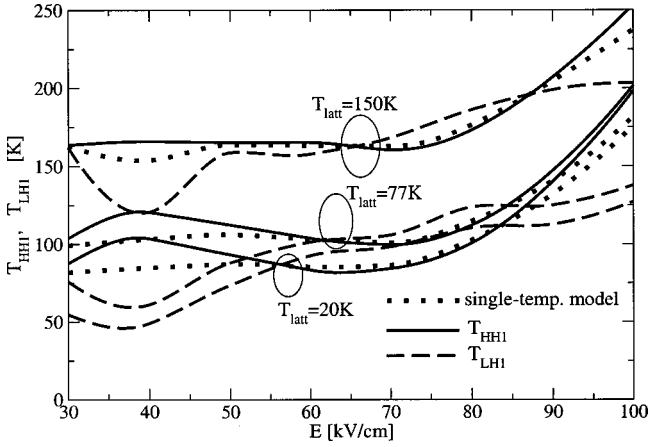

FIG. 3. The bias dependence of the HH1 and LH1 subband temperatures in the *p*-SiGe cascade described in Fig. 1, calculated within the single-temperature or multiple-temperature models, but without hole-hole scattering, at different values of the lattice temperature T_{latt}.

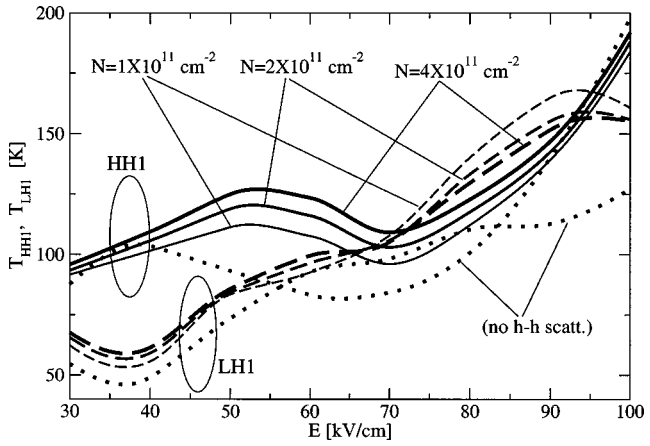

FIG. 5. The bias dependence of the HH1 and LH1 subband temperatures in the *p*-SiGe cascade described in Fig. 1, calculated within the multiple-temperature model, for different values of hole sheet density per period, at $T_{latt}=20$ K.

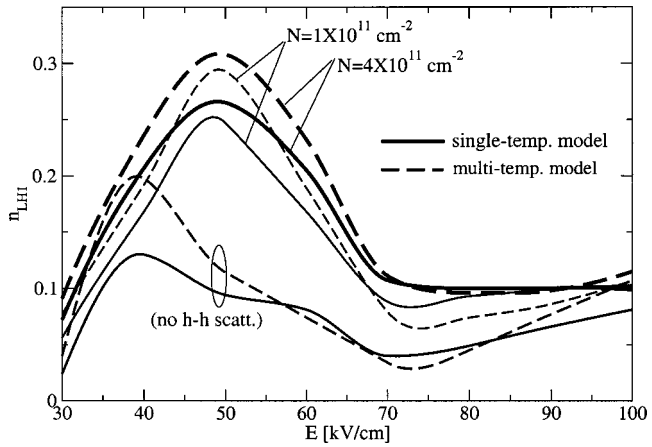

FIG. 6. The bias dependence of the LH1 subband population in the *p*-SiGe cascade described in Fig. 1, calculated within the multiple-temperature model, for different values of hole sheet density per period, at $T_{latt}=20$ K.

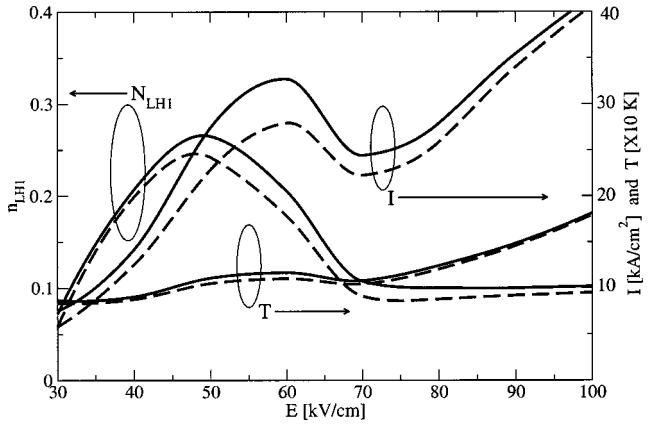

FIG. 8. A comparison of population, temperature, and the current density in the *p*-SiGe cascade described in Fig. 1, calculated within the single-temperature model, for the sheet hole density of 4×10^{11} cm^{-2}, using either the full set of hole-hole scattering processes (solid lines) or the restricted set described in the text (dashed lines).

Figs. 4 and 5. The carriers temperature generally increases with increased hole density, but the dependence is relatively small. The hole temperatures in the two subbands may be quite different for some values of bias, or similar for other values. In a conduction subband cascade this would hardly have any effect on the optical spectra, because of the identical dispersions of subbands and the wave vector conservation in optical transitions, but may be important in transport. On the other hand, in *p*-type cascades the different values of hole temperatures are important for intersubband optical properties as well, because these depend on the hole distributions in the strongly and *unequally* dispersive subbands.

The population of the LH1 subband, and the current density are shown in Figs. 6 and 7. Hole-hole scattering clearly increases both quantities above the values which would hold without this mechanism. Yet, throughout the range of parameters explored in these calculations, the LH1 population remains smaller than that of the HH1 state, implying the existence of population inversion on the interwell HH1$\rightarrow$LH1 optical transition. There is a region (60–70 kV/cm) of negative differential resistance (NDR), and this is more prominent for larger hole densities, because of the increasing role of hole-hole scattering. NDR appears as the interplay of the bias-dependent rate of the interwell processes of this type (which favor small spacing between subbands), and the bias-dependent asymmetry of right $\rightarrow$ left and left $\rightarrow$ right transfers, which requires somewhat larger subband spacings. Although the peak/valley current ratio may not be very large, there exists the possibility of domain formation in the cascade, which would locally shift the energy of the lasing transition and adversely influence the gain.

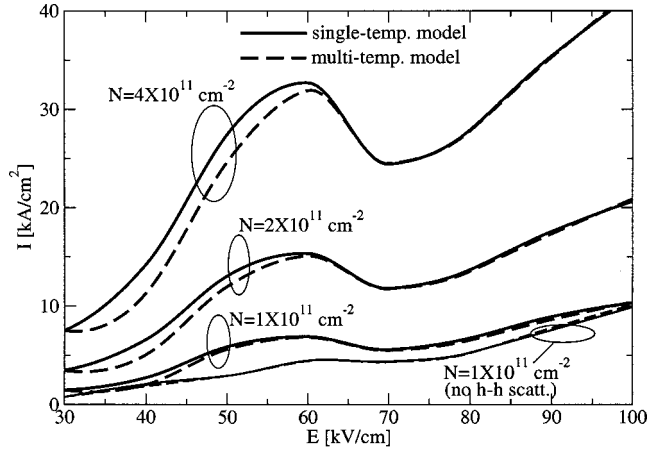

FIG. 7. The bias dependence of the current density in the *p*-SiGe cascade described in Fig. 1, calculated within the multiple-temperature model, for different values of hole sheet density per period, at $T_{latt}=20$ K.

With two initial and two final states, the number of different hole-hole scattering processes is clearly much larger than the number of single-hole processes. As the number of states per period increases it soon becomes impossible to account for all of them in any realistic computation time. It is therefore necessary to limit the number of processes by a suitable selection criterion. A possible simple choice that we have explored here (given that the structure has just two subbands localized in a single well) is to retain only those processes which involve just two different states, i.e., the subscripts in any particular scattering rate $w_{i,j,f,g}$ may take only two values, but these will of course run over all states in the system. This greatly reduces the number of processes included, e.g., in a system with six states per period the saving is almost by a factor of 20. The scattering matrix elements generally decrease when spatially more displaced states are involved, indicating that this might be a good approximation. On the other hand, the number of processes discarded by such a restriction is usually much larger than the number of processes retained, which makes the approximation questionable. This is best resolved by direct calculation.

A cascade with two states per period is a good testing grounds for the quality of the above approximation, because the total number of processes is just three times larger than the number in the restricted set, so both calculations are manageable in real time. Results obtained in case of a large hole density, where hole-hole scattering is relatively more important, are compared in Fig. 8. Using the restricted set of hole-hole scattering processes leads to some underestimate of the

current, and also of the LH1 state population, by no more than 20%. The errors in the calculated hole temperature are smaller. These results are obtained within the single-temperature model, but the comparison is quite similar for multiple-temperature model. Therefore, with quite acceptable overall accuracy, the above process selection rule was adopted in calculations of more complex structures. While the number of discarded processes in such cases will be far larger than those retained, it is also true that most of the former include states which are more remote than are the states in the simple cascade, and so their contribution to hole-hole scattering is negligible. Therefore, we believe that using the restricted set of processes in complex structures generally provides a similar level of accuracy as in simple cascades. Yet another selection criterion might be based on the energy of the states, because carrier-carrier scattering favors small energy differences. This is less safe to employ for holes than for electrons, because hole subbands may greatly change their energy spacing as the in-plane wave vector varies, and such additional selection was not implemented in this work. Certainly, in structures which have more than two subbands localized in a single well there may exist efficient hole-hole processes which include, for example, three different subbands,[39] and the above selection criterion should then be appropriately expanded.

Next we consider two Si/SiGe quantum cascade structures which are more complex, having three wells per period, in order to explore the role of optical phonon scattering in cooling of holes. It has been argued[13] that electron cooling in the injector portion of AlGaAs based QCLs requires spacing between subsequent subbands to be somewhat less than the optical phonon energy. In such a case only those electrons in the upper subband which are hot, i.e., have sufficient in-plane kinetic energy, will be able to emit an optical phonon (by the fast, Frölich polar interaction) and terminate in the lower subband with a small kinetic energy. However, there is no polar optical phonon scattering in SiGe, and the nonpolar deformation potential scattering differs considerably in its dependence on the initial carrier energy. Although the Ge-Ge mode phonon energy is almost the same as that for LO phonons in GaAs, nonpolar optical phonon scattering is still slow when the subband spacing is equal to or slightly greater than the phonon energy (whereas the polar optical phonon scattering rate peaks in this case). Only at substantially larger subband spacings does nonpolar phonon scattering become a very fast process.

Cascade *A* has the structure: 9 ML $Ge_{0.36}Si_{0.64}$/6 ML Si/16 ML $Ge_{0.3}Si_{0.7}$/6ML Si/14ML $Ge_{0.3}Si_{0.7}$/6ML Si (ML =crystalline monolayer, ≈2.75 Å), and is grown on $Ge_{0.2}Si_{0.8}$ virtual substrate. We will refer to these three wells as No. I, No. II, and No. III, respectively. Each of the wells has two low-lying states (HH1 and LH1), spaced by 36 meV in well No. I, and by 30 meV in wells No. II and No. III. At a bias of 60 kV/cm the HH1 state from any preceeding well is almost aligned with the LH1 state of the next well, as shown in Fig. 9. The aligned subbands are localized in individual wells at $\mathbf{k}_{\|}=0$, because HH and LH show (almost) no interaction at the zone center, but for any small finite $\mathbf{k}_{\|}$ they become strongly hybridized, and scattering between them

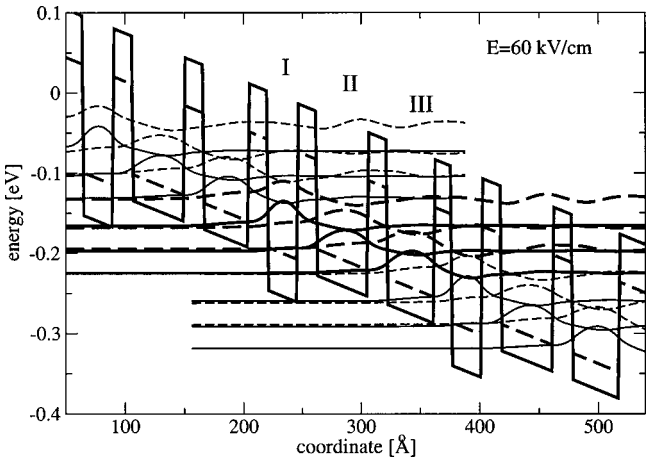

FIG. 9. The potential profile in cascade structure *A*, described in the text, and the wave functions calculated at $\mathbf{k}_{\|}=0$. Wave functions assigned to the central period are shown in thick, and their replicas in the adjacent two periods in thin lines (solid—heavy hole, dashed—light hole states).

should be fast and with little carrier heating. Hence, it is more meaningful to talk about the average temperature of the hybridized pairs of states, than of individual states. Using the same reasoning as for electrons in a GaAs/AlGaAs cascade one would expect that the hole temperature decreases when going from well No. I (where the optical phonon transitions generate no carrier heating) into No. II and then No. III (where these transitions should cool the holes). However, calculations (performed at $T_{\text{latt}}=20$ K and hole sheet density of 10^{11} cm^{-2}) show that quite the opposite happens: the temperature of the hybridized pair [HH1 No. III +LH1 No. I (R)] is 261 K, that of (HH1 No. II +LH1 No. III) is 211 K, and that of (HH1 No. I +LH1 No. II) is 126 K. Although at these temperatures many holes have sufficient kinetic energy to compensate for the 6 meV deficiency and emit an optical phonon (and those which do so will indeed cool), the process is still slow to make a significant influence on overall temperatures. The transport between wells is here mostly due to acoustic phonons or fully elastic processes (alloy and hole-hole scattering), and these induce only heating of holes, while the role of cooling is left to intrasubband acoustic phonon scattering.

In another cascade (*B*, Fig. 10), that has the structure: 30 ML $Ge_{0.3}Si_{0.7}$/6ML Si/13ML $Ge_{0.35}Si_{0.65}$/6ML Si/14ML $Ge_{0.35}Si_{0.65}$/6ML Si, also grown on a $Ge_{0.2}Si_{0.8}$ virtual substrate, the HH1-LH1 spacing is 30 meV in the well No. I, and by 38 meV in wells No. II and No. III. Similarly as in cascade *A*, the alignment of HH1 and LH1 states from subsequent wells here occurs at a bias of 55 kV/cm. The temperatures of the [HH1 No. III +LH1 No. I (R)], (HH1 No. II +LH1 No. III), and (HH1 No. I +LH1 No. II) pairs of states are now calculated to be 122, 183, and 210 K, respectively. The generally lower temperatures in this case partly result from a smaller current (twice smaller than in cascade *A*), but it is interesting to note that the temperature gradient is now reversed. Therefore, although in this case *all* holes in wells No. II and No. III can emit optical phonons, which results in their increased kinetic energy in the lower subband, it still appears advantageous to allow for some

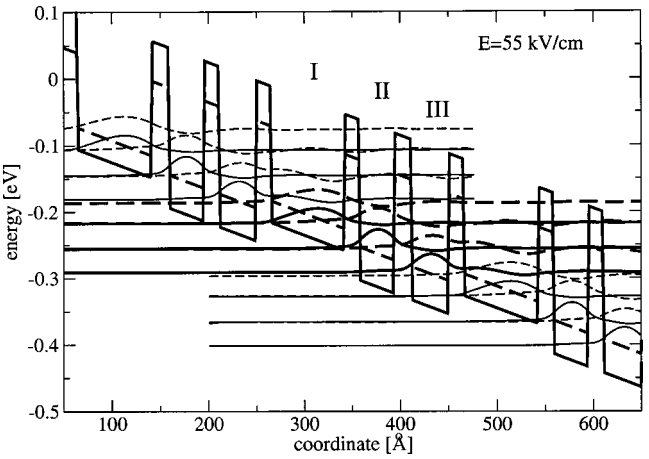

FIG. 10. Same as Fig. 9, but for cascade structure *B*.

amount of interwell optical phonon scattering to replace other processes which are almost or fully elastic, and hence produce much more heating. However, for much larger HH -LH spacing the optical phonon scattering would become a dominant process and, being very fast, it would only lead to both large current density and heating.

IV. CONCLUSION

A self-consistent energy balance method for the simulation of hole transport in *p*-Si/SiGe quantum cascade structures has been developed and used to study carrier distributions and carrier heating effects. The scattering mechanisms included in the model are the alloy disorder, acoustic and optical phonon scattering, and hole-hole scattering accounting for the in-plane anisotropy of both the hole subband structure and the scattering rates. In simple cascade structures with just one HH and one LH subband per period the HH and LH temperatures can differ by up to a factor of 2, although the weighted average is close to the value obtained from a single carriers temperature simulation. In more complex structures, the carrier temperatures of subbands in adjacent wells differ markedly, and we have shown how such a device can be designed to optimize carrier cooling via optical phonon scattering. The effect of hole-hole scattering on carrier dynamics and carrier heating has also been investigated. Hole-hole scattering results in increased carrier temperatures in a simple (superlattice) cascade, and in much stronger coupling between states in adjacent wells. This leads to substantially higher currents, and more pronounced NDR features in the current-voltage response.

ACKNOWLEDGMENTS

This work is supported by DARPA /USAF Contract No. F19628-99-C-0074. The authors thank R. A. Soref (Hanscom AFB for useful discussions.

[1]R. A. Soref, L. Friedman, and G. Sun, Superlattices Microstruct. **23**, 427 (1998).
[2]L. Friedman, R. A. Soref, G. Sun, and Y. Lu, IEEE J. Sel. Top. Quantum Electron. **4**, 1029 (1998).
[3]G. Dehlinger, L. Diehl, U. Gennser, H. Sigg, J. Faist, K. Ensslin, D. Grutzmacher, and E. Muller, Science **290**, 2277 (2000).
[4]L. Diehl *et al.*, Appl. Phys. Lett. **81**, 4700 (2002).
[5]S. A. Lynch *et al.*, Appl. Phys. Lett. **81**, 1543 (2002).
[6]D. J. Paul *et al.*, Physica E (Amsterdam) **16**, 147 (2003).
[7]R. Bates *et al.*, Appl. Phys. Lett. **83**, 4092 (2003).
[8]K. Donovan, P. Harrison, and R. W. Kelsall, J. Appl. Phys. **89**, 3084 (2001).
[9]K. Kalna, C. Y. L. Cheung, and K. A. Shore, J. Appl. Phys. **89**, 2001 (2001).
[10]D. Indjin, P. Harrison, R. W. Kelsall, and Z. Ikonić, J. Appl. Phys. **91**, 9019 (2002).
[11]D. Indjin, P. Harrison, R. W. Kelsall, and Z. Ikonić, Appl. Phys. Lett. **81**, 400 (2002).
[12]D. Indjin, P. Harrison, R. W. Kelsall, and Z. Ikonić, IEEE Photonics Technol. Lett. **15**, 15 (2003).
[13]P. Harrison, D. Indjin, and R. W. Kelsall, J. Appl. Phys. **92**, 6921 (2002).
[14]R. C. Iotti and F. Rossi, Appl. Phys. Lett. **78**, 2902 (2001).
[15]R. C. Iotti and F. Rossi, Phys. Rev. Lett. **87**, 146603 (2001).
[16]H. Callebaut, S. Kumar, B. S. Williams, Q. Hu, and J. L. Reno, Appl. Phys. Lett. **83**, 207 (2003).
[17]R. Oberhuber, G. Zandler, and P. Vogl, Phys. Rev. B **58**, 9941 (1998).
[18]M. V. Fischetti, Z. Ren, P. M. Solomon, M. Yang, and K. Rim, J. Appl. Phys. **94**, 1079 (2003).
[19]Z. Ikonić, R. W. Kelsall, and P. Harrison, Phys. Rev. B **69**, 235308 (2004).
[20]Z. Ikonić, P. Harrison, and R. W. Kelsall, Phys. Rev. B **64**, 245311 (2001).
[21]B. A. Foreman, Phys. Rev. B **48**, 4964 (1993).
[22]Z. Ikonić, P. Harrison, and R. W. Kelsall, Phys. Rev. B **64**, 125308 (2001).
[23]M. El kurdi, G. Fishman, S. Sauvage, and P. Boucaud, Phys. Rev. B **68**, 165333 (2003).
[24]K. Bhaumik, B. K. Ridley, and Y. Shacham-Diamand, J. Appl. Phys. **74**, 5546 (1993).
[25]F. L. Madarasz and F. Szmulowicz, Phys. Rev. B **24**, 4611 (1981).
[26]F. Szmulowicz, Phys. Rev. B **28**, 5943 (1983).
[27]J. M. Hinckley, and J. Singh, J. Appl. Phys. **76**, 4192 (1994).
[28]K. Reimann, R. A. Kaindl, and M. Woerner, Phys. Rev. B **65**, 045302 (2001).
[29]K. Greipel and U. Rössler, Semicond. Sci. Technol. **7**, 487 (1992).
[30]D. C. Look, D. K. Lorance, J. R. Sizelove, C. B. Stutz, K. R. Evans, and D. W. Whitson, J. Appl. Phys. **71**, 260 (1992).
[31]K. Yeom, J. M. Hinckley, and J. Singh, J. Appl. Phys. **80**, 6773 (1996).
[32]M. J. Kearney and A. I. Horrell, Semicond. Sci. Technol. **13**, 174 (1998).
[33]G. Gilat and N. R. Bharatiya, Phys. Rev. B **12**, 3479 (1975).
[34]S.-Y. Ren and W. A. Harrison, Phys. Rev. B **23**, 762 (1981).
[35]S. I. Kurganskii, O. I. Dubrovskii, and E. P. Domashevskaya, Phys. Status Solidi B **129**, 293 (1985).
[36]H. Eschrig, in *Optimized LCAO-method* (Springer, Berlin, 1989), p. 113–219 Chap. 6.
[37]P. Harrison, *Quantum Wells, Wires and Dots: Theoretical and Computational Physics* (Wiley, Chichester, 1999).
[38]J. H. Smet, C. G. Fonstad, and Q. Hu, J. Appl. Phys. **79**, 9305 (1996).
[39]P. Kinsler, P. Harrison, and R. W. Kelsall, Phys. Rev. B **58**, 4771 (1998).
[40]S. C. Lee and I. Galbraigth, Phys. Rev. B **59**, 15796 (1999).
[41]S. C. Lee and I. Galbraith, Physica B **272**, 237 (1999).
[42]C. J. Williams, E. Corbin, M. Jaros, and D. C. Herbert, Physica B **254**, 240 (1998).
[43]D. M. Newns, and P. C. Pattnaik, Phys. Rev. B **54**, 16313 (1996).
[44]D. K. Ferry, S. M. Goodnick, and K. Hess, Physica B **272**, 538 (1999).
[45]A. Kahan, M. Chi, and L. Friedman, J. Appl. Phys. **75**, 8012 (1994).
[46]C. G. Van de Walle and R. M. Martin, Phys. Rev. B **34**, 5621 (1986).
[47]P. Murzyn *et al.*, Appl. Phys. Lett. **80**, 1456 (2002).