Full band modeling of the excess current in a delta-doped silicon tunnel diode

Cristian Rivas[a)] and Roger Lake[b)]
Department of Electrical Engineering, University of California, Riverside, California 92521-0204

William R. Frensley
Eric Jonsson School of Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688

Gerhard Klimeck
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109

Phillip E. Thompson and Karl D. Hobart
Naval Research Laboratory, Code 6812, Washington, D.C. 20375-5347

Sean L. Rommel
Microelectronics Engineering Department, Rochester Institute of Technology, Rochester, New York 14263-5604

Paul R. Berger[c)]
Department of Electrical Engineering, The Ohio State University, Columbus, Ohio 43210-1272

The current of a molecular beam epitaxially grown Sb and B delta-doped Si tunnel diode is simulated in all regions of tunneling: peak, valley, and post-valley turn-on. All three regions of the $I–V$ are qualitatively captured by the calculations. The inclusion in the model of bandtail states gives rise to the excess current and the post-valley turn on of the tunnel current. This excess current is dominated by the direct coherent tunneling component of the current tunneling from gap state to gap state. The crossover between phonon-assisted and direct occurs immediately after the valley minimum. The calculated voltages quantitatively match the experimental measurements. The magnitude of the calculated current is approximately a factor of 5.4 too small. Sources of error are analyzed. The current calculations use a second neighbor sp^3s* planar orbital basis within the nonequilibrium Green function formalism. © *2003 American Institute of Physics.*
[DOI: 10.1063/1.1606114]

I. INTRODUCTION

Recent demonstrations of silicon based tunnel diodes fabricated with low-temperature molecular beam epitaxy (LT-MBE)[1–12] have exhibited a maximum current density of $150 \, kA/cm^2$,[13,14] and maximum peak to valley current ratio (PVCR) of 6.[10] These demonstrations have important electronic device implications. The fabrication process is compatible with the complementary metal-oxide-semiconductor (CMOS) or Si/Si_xGe_{1-x} bipolar technology. The current density and PVCR are sufficient for high speed switching applications. The process is also interesting from a device physics perspective since, in contrast to the alloy construction techniques of the 1960's, the LT-MBE process allows one to engineer the junction potential for investigation of its effect on the current and peak-to-valley current ratio.[3,6–8]

The peak-to-valley current ratio, one figure of merit for these devices, is limited by the excess current. The theory of the excess current developed by Chynoweth, Feldman, and Logan[15] relates the excess current to the density of states in the band gap. We incorporate that theory into a full band quantum model of the tunnel diode. It was an open question as to whether the inclusion of a smooth distribution of states in the gap would give rise in our simulations to the observed qualitative features in the experimentally measured current–voltage response of our LT-MBE grown devices. We find that the inclusion of gap states in the contact regions results in the correct trends in both the valley current and the post-valley turn-on of the tunnel current. Furthermore, we find that the excess current is dominated by direct tunneling from gap state to gap state whereas the peak current is determined by phonon assisted tunneling.

The theory of interband tunneling in Si tunnel diodes developed from the analytical work of the late 1950's and early 1960's[16–19] and came to an end by the middle 1970's.[20] The problem was recently revisited to calculate enhanced direct tunneling,[21] single electron tunneling,[22] and the temperature dependence of the excess current.[23] We return to this problem applying the numerical techniques of full-band quantum device modeling developed during the last two decades.[24–29] The full band techniques were formulated to model coherent tunneling;[24–28] or if they were formulated to include phonon scattering, they were implemented in a single band model.[29] Full band quantum transport techniques were

---
[a)]Also at Eric Jonsson School of Engineering, University of Texas at Dallas, Richardson, TX 75083-0688.
[b)]Electronic mail: rlake@ee.ucr.edu
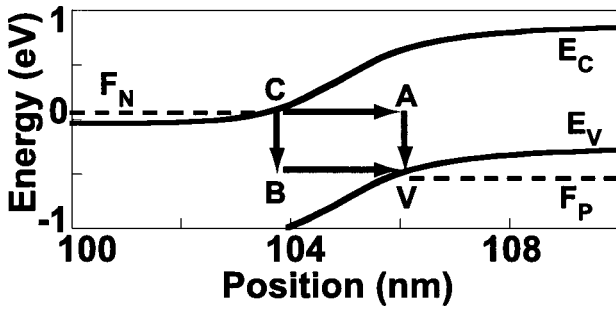[c)]Also at Department of Physics, The Ohio State University, Colombus, OH 43210-1106.

FIG. 1. Band profile for the tunnel diode biased at 0.6 V (at the turn-on of the post-valley tunnel current) illustrating the excess current tunneling paths.

combined with the formalism of Caroli et al.[30] to calculate the indirect phonon-assisted tunneling current in a Si tunnel diode.[31] We now describe that theory and approach in detail, and we enhance it to model all regions of the tunnel current: peak current, valley current, and post-valley turn-on current.

The particular device that we model is the well-characterized tunnel diode described by Thompson et al.[4] Figure 1 shows the band diagram near the tunnel barrier calculated from the SIMS doping profile measured after a 700 °C rapid thermal anneal assuming complete dopant activation.[32] Modeling the PVCR in these devices requires calculating both the peak current and the excess current. The peak current is an indirect (X-Γ), interband, phonon-assisted process. The mechanism that we model for the excess current was described and analyzed by Chynoweth, Feldman, and Logan.[15] Electrons tunnel from gap states labeled "B" in Fig. 1 to the valence band or from the conduction band to gap states labeled "A." In this model, the gap states signified by "B" are a smooth tail of donor levels lying below the conduction band edge and the states signified by "A" are a smooth tail of acceptor levels lying above the valence band edge. The rate limiting mechanism is the interband tunneling. Thus, the tail states are equilibrated with their respective Fermi levels; "B" with the conduction band Fermi level, and "A" with the valence band Fermi level.

II. APPROACH

Our approach uses the nonequilibrium Green function formalism[29–31] in both a second neighbor sp^3s^{*}[33] and a nearest neighbor $sp^3s^{*}d^5$ planar orbital basis.[34] The Hamiltonian matrix elements are optimized using a genetic algorithm.[35]

The derivation of the phonon assisted tunneling current follows closely that of Caroli et al. leading to their Eq. (49).[30] We start with the general expression for the current,

$$J = \frac{ie}{A\hbar} \sum_{\mathbf{k}} \int \frac{dE}{2\pi} \text{tr}\{\Gamma^{\mathcal{L}}[f^{\mathcal{L}}G^{>} + (1-f^{\mathcal{L}})G^{<}]\} \quad (1)$$

and use the following relations for $G^{<}$ and $G^{>}$:

$$G^{>} = -i(1-f^{\mathcal{L}})G^R\Gamma^{\mathcal{L}}G^A - i(1-f^{\mathcal{R}})G^R\Gamma^{\mathcal{R}}G^A$$
$$+ G^R\Sigma^{>}G^A, \quad (2)$$

$$G^{<} = if^{\mathcal{L}}G^R\Gamma^{\mathcal{L}}G^A + if^{\mathcal{R}}G^R\Gamma^{\mathcal{R}}G^A + G^R\Sigma^{<}G^A. \quad (3)$$

All of our notation corresponds to that defined in Ref. 29. In Eqs. (1)–(3), the superscripts $\mathcal{L}$ and $\mathcal{R}$ indicate that the

quantity is injected from the left or right contact, respectively. Γ is obtained from the surface Green functions of the contact regions.[29] G^R and G^A are the retarded and advanced Green functions, respectively. tr{ } indicates a trace over the 20 orbitals in the $sp^3 s^{*}$ basis or 40 orbitals in the $sp^3s^{*}d^5$ basis. $\mathbf{k}$ is the transverse, two-dimensional wave vector in the $x-y$ plane. A is the cross-sectional area, and e is the electron charge, -1.602×10^{-19} C. $f^{\mathcal{L}}$ and $f^{\mathcal{R}}$ are the Fermi factors of the left and right contacts, respectively. Working to second order in the electron phonon coupling, we use the bare Green functions denoted with a lower case g, undressed by the electron phonon coupling, and, substituting Eqs. (2) and (3) into Eq. (1) obtain

$$J = \frac{e}{A\hbar} \sum_{\mathbf{k}} \int \frac{dE}{2\pi} \text{tr}\{\Gamma^{\mathcal{L}}g^R\Gamma^{\mathcal{R}}g^A\}(f^{\mathcal{L}} - f^{\mathcal{R}})$$
$$+ i\text{tr}\{\Gamma^{\mathcal{L}}g^R[f^{\mathcal{L}}\sigma^{>} + (1-f^{\mathcal{L}})\sigma^{<}]g^A\}, \quad (4)$$

where σ denotes that the self energy is also calculated using the bare Green functions. The first term of Eq. (4) is the coherent tunneling, and the second term is the phonon assisted tunneling current. For the device under consideration, we find that the peak current is determined by the phonon assisted current and the excess current after the valley is determined by the coherent tunneling current. The derivation of the coherent tunneling term is finished, and we now focus on the phonon-assisted tunneling term.

We begin with the intervalley deformation potential Hamiltonian for the electron phonon interaction

$$H' = \sum_{\mathbf{q}} \sum_{\nu} \left[\frac{\hbar}{2\rho\omega_{\mathbf{q},\nu}V} \right]^{1/2} (a_{\mathbf{q},\nu} + a^{\dagger}_{-\mathbf{q},\nu})e^{i\mathbf{q}\cdot\mathbf{r}}(D_t K), \quad (5)$$

where $\mathbf{q}$ is the three-dimensional phonon wave vector, ν is the polarization, ρ is the Si density, 2.328 g/cm^3, V is the volume, and ω is the frequency. D_t is the deformation potential, and K is the transverse wave vector coupling a transverse Δ valley to the Γ valley, e.g., $0.85\frac{2\pi}{a}(100)$ where a is the lattice constant of 5.431 Å. We derive the second order self energies for dispersionless phonons as described in Appendix A of Ref. 29 evaluating the matrix elements as described by Boykin[36] to obtain

$$\sigma^{<}_{\alpha_1,n;\alpha_2,n}(E) = \frac{4}{aA}|U|^2 \zeta_{\alpha_1,\alpha_2}$$
$$\times \sum_{\mathbf{k}} \{g^{<}_{\alpha_1,n;\alpha_2,n}(\mathbf{k};E+\hbar\omega)[n_B(\omega)+1]$$
$$+ g^{<}_{\alpha_1,n;\alpha_2,n}(\mathbf{k};E-\hbar\omega)n_B(\omega)\}, \quad (6)$$

and

$$\sigma^{>}_{\alpha_1,n;\alpha_2,n}(E) = \frac{4}{aA}|U|^2 \zeta_{\alpha_1,\alpha_2}$$
$$\times \sum_{\mathbf{k}} \{g^{>}_{\alpha_1,n;\alpha_2,n}(\mathbf{k};E-\hbar\omega)[n_B(\omega)+1]$$
$$+ g^{>}_{\alpha_1,n;\alpha_2,n}(\mathbf{k};E+\hbar\omega)n_B(\omega)\}, \quad (7)$$

where A is the cross sectional area, $|U|^2 = \hbar(D_t K)^2/(2\rho\omega)$, n labels the monolayer consisting of two atomic planes, E is the energy, $n_B(\hbar\omega)$ is the Bose–Einstein

factor, α is a localized orbital, and $\zeta_{\alpha_1,\alpha_2}$ equals 1.0 when the two orbitals α_1 and α_2 lie in the same atomic plane and zero otherwise. The sum over polarization has been replaced by a factor of 2. The sum over polarization was ignored in our previous calculations.[31] Now, we expand out $g^<$ and $g^>$ in Eqs. (6) and (7) using the first two terms of Eqs. (2) and (3). For compactness, we drop the orbital indices and define the diagonal blocks of the left- and right-injected spectral functions, respectively, as

$$a_n^{\mathcal{L}}=g_{n,1}^R\Gamma_{1,1}g_{1,n}^A=g_{n,1}^A\Gamma_{1,1}g_{1,n}^R,\tag{8}$$

$$a_n^{\mathcal{R}}=g_{n,N}^R\Gamma_{N,N}g_{N,n}^A=g_{n,N}^A\Gamma_{N,N}g_{N,n}^R,\tag{9}$$

and then, suppressing all subscripts, the self energies become

$$\begin{aligned}\sigma^>(E)=\frac{4|U|^2\zeta}{aA}\sum_{\mathbf{k}}\ &\{[1-f^{\mathcal{L}}(E-\hbar\omega)]a^{\mathcal{L}}(\mathbf{k};E-\hbar\omega)[n_B(\hbar\omega)+1]\\ &+[1-f^{\mathcal{R}}(E-\hbar\omega)]a^{\mathcal{R}}(\mathbf{k};E-\hbar\omega)[n_B(\hbar\omega)+1]\\ &+[1-f^{\mathcal{L}}(E+\hbar\omega)]a^{\mathcal{L}}(\mathbf{k};E+\hbar\omega)n_B(\hbar\omega)\\ &+[1-f^{\mathcal{R}}(E+\hbar\omega)]a^{\mathcal{R}}(\mathbf{k};E+\hbar\omega)n_B(\hbar\omega)\},\end{aligned}\tag{10}$$

and

$$\begin{aligned}\sigma^<(E)=\frac{4|U|^2\zeta}{aA}\sum_{\mathbf{k}}\ &\{[f^{\mathcal{L}}(E-\hbar\omega)]a^{\mathcal{L}}(\mathbf{k};E-\hbar\omega)n_B(\hbar\omega)\\ &+f^{\mathcal{R}}(E-\hbar\omega)a^{\mathcal{R}}(\mathbf{k};E-\hbar\omega)n_B(\hbar\omega)\\ &+f^{\mathcal{L}}(E+\hbar\omega)a^{\mathcal{L}}(\mathbf{k};E+\hbar\omega)[n_B(\hbar\omega)+1]\\ &+f^{\mathcal{R}}(E+\hbar\omega)a^{\mathcal{R}}(\mathbf{k};E+\hbar\omega)[n_B(\hbar\omega)+1]\}.\end{aligned}\tag{11}$$

Substituting these expressions for $\sigma^<$ and $\sigma^>$ back into the second term of Eq. (4), we obtain

$$\begin{aligned}J_1=\frac{2e(D_tK)^2}{\rho\omega a}\int\frac{d^2\mathbf{k}_{\mathcal{L}}}{4\pi^2}\int\frac{d^2\mathbf{k}_{\mathcal{R}}}{4\pi^2}\int\frac{dE}{2\pi}\sum_n\ &\mathrm{tr}\{a_n^{\mathcal{L}}(\mathbf{k},E)\\ \times[&f^{\mathcal{L}}(E)(1-f^{\mathcal{R}}(E-\hbar\omega))a_n^{\mathcal{R}}(\mathbf{k}_{\mathcal{R}},E-\hbar\omega)(n_B(\hbar\omega)+1)\\ +&f^{\mathcal{L}}(E)(1-f^{\mathcal{R}}(E+\hbar\omega))a_n^{\mathcal{R}}(\mathbf{k}_{\mathcal{R}},E+\hbar\omega)n_B(\hbar\omega)\\ -&(1-f^{\mathcal{L}}(E))f^{\mathcal{R}}(E-\hbar\omega)a_n^{\mathcal{R}}(\mathbf{k}_{\mathcal{R}},E-\hbar\omega)n_B(\hbar\omega)\\ -&(1-f^{\mathcal{L}}(E))f^{\mathcal{R}}(E+\hbar\omega)a_n^{\mathcal{R}}(\mathbf{k}_{\mathcal{R}},E+\hbar\omega)(n_B(\hbar\omega)+1)]\},\end{aligned}\tag{12}$$

where we have used the cyclic invariance of the trace and relation (8). In our device configuration as shown in Fig. 1, the left injected spectral function is injected from one of the four equivalent transverse Δ valleys of the conduction band and the right injected spectral function is injected from the Γ valley of the valence band. To emphasize this, we will replace the superscripts $\mathcal{L}$ and $\mathcal{R}$ by Δ and Γ, respectively, e.g., $a^{\mathcal{L}}$ and $a^{\mathcal{R}}$ become a^{Δ} and a^{Γ}, respectively.

With this substitution, the trace in Eq. (12) results in terms such as

$$\mathrm{tr}\{a_n^{\Delta}(\mathbf{k}_{\Delta},E)a_n^{\Gamma}(\mathbf{k}_{\Gamma},E-\hbar\omega)\},\tag{13}$$

where Eq. (13) is the overlap of the underlying Bloch states and is very small. However, we know that the short wavelength transverse acoustic and transverse optical phonons distort the unit cell resulting in nonzero matrix elements between the Δ and Γ states. We therefore include these matrix elements, $|\langle\Gamma|H_{TA}|\Delta\rangle|^2$ and $|\langle\Gamma|H_{TO}|\Delta\rangle|^2$ in the strength D_tK and trace over each spectral function individually. Our final expression for the phonon assisted tunneling current is then

$$J_1 = \frac{4e(D_t K)^2}{\rho \omega a} \int_\Delta \frac{d^2 \mathbf{k}_\Delta}{4\pi^2} \int_\Gamma \frac{d^2 \mathbf{k}_\Gamma}{4\pi^2} \int \frac{dE}{2\pi} \sum_n \left[\mathrm{tr}\{a_n^\Delta(\mathbf{k}_\Delta, E)\} \right.$$

$$\times [\mathrm{tr}\{a_n^\Gamma(\mathbf{k}_\Gamma, E - \hbar\omega)\} f^\Delta(E)(1 - f^\Gamma(E - \hbar\omega))(n_B(\hbar\omega) + 1)$$

$$+ \mathrm{tr}\{a_n^\Gamma(\mathbf{k}_\Gamma, E + \hbar\omega)\} n_B(\hbar\omega) f^\Delta(E)(1 - f^\Gamma(E + \hbar\omega))$$

$$- \mathrm{tr}\{a_n^\Gamma(\mathbf{k}_\Gamma, E - \hbar\omega)\}(1 - f^\Delta(E)) f^\Gamma(E - \hbar\omega) n_B(\hbar\omega)$$

$$\left. - \mathrm{tr}\{a_n^\Gamma(\mathbf{k}_\Gamma, E + \hbar\omega)\}(1 - f^\Delta(E)) f^\Gamma(E + \hbar\omega)(n_B(\hbar\omega) + 1)]\right]. \tag{14}$$

In Eq. (14), the integral $\int_\Delta d^2 \mathbf{k}_\Delta / (4\pi^2)$ is performed around the point $0.85 \frac{2\pi}{a}(100)$ for one of the four equivalent conduction band valleys. Since there are four equivalent valleys, we multiply in a factor of 4 in going from Eq. (12) to (14). The integral $\int_\Gamma d^2 \mathbf{k}_\Gamma / (4\pi^2)$ is performed around the Γ point. In breaking up the $\mathrm{tr}\{a^\Delta a^\Gamma\}$ into $\mathrm{tr}\{a^\Delta\}\mathrm{tr}\{a^\Gamma\}$ we introduce an extra summation over spin, and, therefore, divide through by a factor of 2 in going from Eq. (12) to (14). Equation (14) is essentially Fermi's Golden Rule written in Green function form. It is a full-band version of Eq. (49) of Caroli *et al.*[30] written for the interband TO and TA phonons.

Equation (14) is evaluated for each type of phonon that contributes to the current and the contributions are added. Two different phonons provide the main contribution to the current, the transverse acoustic (TA) phonon with an energy of 18.4 meV and the transverse optical (TO) phonon with an energy of 57.6 meV.[37] Previously, using the simplest deformation potential concepts, we made an order of magnitude estimate for the electron–phonon matrix element of $\Xi_d + \Xi_u/3 - a = 2.5$ eV using values from Ref. 38. For a phonon wave vector of $0.85\ 2\pi/a$, we obtained an average intervalley deformation potential for the interband (100) TA phonon of $D_t K = 2.45 \times 10^8$ eV/cm. There exists no similar method to estimate the matrix element for the TO phonon. However, we know experimentally that the contribution to the total current from the TA and TO phonons is almost equal.[39] We used this fact to estimate a value of $D_t K$ for the TO phonon by comparing the two numerically calculated

current contributions. We estimated the value to be 5.6×10^8 eV/cm.

The difference between the calculations described herein and the calculations described previously[31] lies in the limits of integration in the energy and momentum integrals in Eq. (14). Previously, the energy integral in Eq. (14) was integrated from the lowest point in the intrinsic Si conduction band to the highest point in the valence band. Once the nominal band edges uncrossed as shown in Fig. 1, the energy integral in Eq. (14) was zero resulting in zero excess current. In this work, the energy integral extends from a minimum of $20\ k_B T$ below the lowest Fermi level (the Fermi level of the valence band) to $20\ k_B T$ above the highest Fermi level (the Fermi level of the conduction band) as illustrated in Fig. 2. The momentum integrals are extended up to a maximum $|\mathbf{k}|$ corresponding to the energy $F_N + 20 k_B T$ for electrons and $F_P - 20 k_B T$ for holes. The maximum k value is used for the single $\mathbf{k}$ integration for the coherent current.

The Green function formalism allows one to naturally include the effect of bandtails in the contacts. The contact regions shown in Fig. 2 are defined as the region to the left of where the electron Fermi level crosses the conduction band edge and the region to the right of where the hole Fermi level crosses the valence band. Including an imaginary potential in these regions, gives rise to a nonzero density of states within the intrinsic band gap. Figure 3 shows the 1D density of states, D_{1D}, in bulk Si for three values of the imaginary potential, 2, 5, and 10 meV, corresponding to
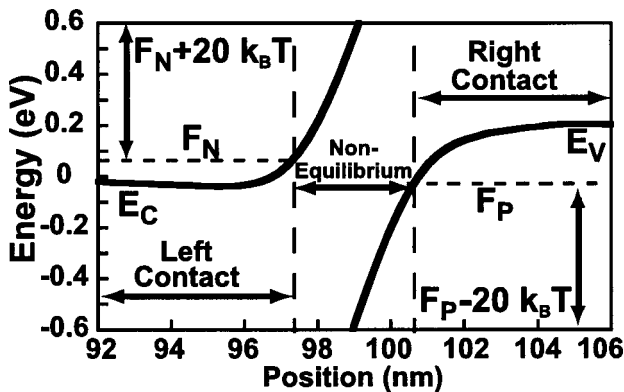

FIG. 2. Closeup of the calculated tunnel junction band profile showing the limits of the energy integration and the spatial partitioning for the contact and nonequilibrium regions.
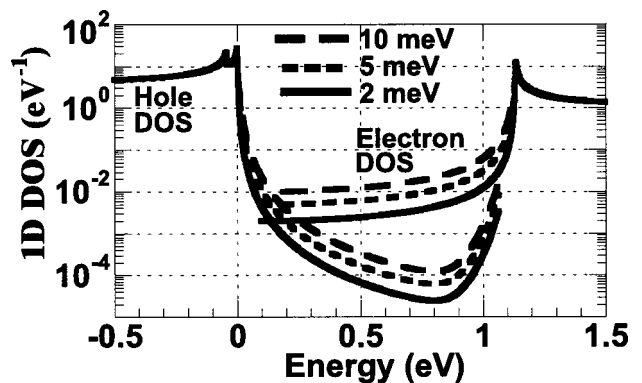

FIG. 3. One dimensional density of states for the electron and hole bands for three different coherent lifetimes: 33, 66, and 165 fs corresponding to the imaginary potentials 10, 5, and 2 meV, respectively.
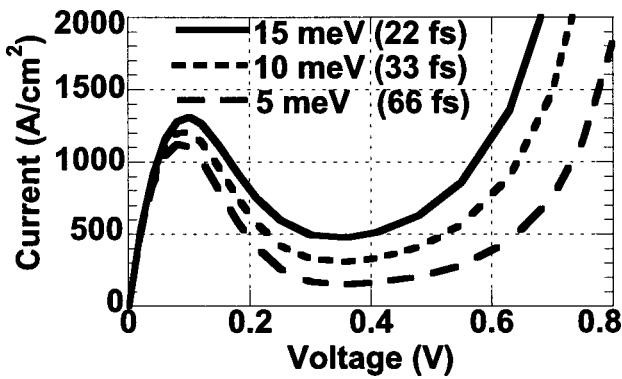
FIG. 4. Calculated I–V's for lifetimes in the contacts corresponding to the imaginary potentials as shown. All current shown, peak, valley, and post-valley is tunnel current.

coherent lifetimes of 165, 66, and 33 fs, respectively, calculated from

$$D_{1D} = -2\mathrm{tr}\{\mathrm{Im}[G^R_{n,n}(\mathbf{k},E)]\}, \qquad (15)$$

where $\mathbf{k}=(k_x,k_y)$ is $(0,0)$ for the valence band calculation and $0.85(2\pi/a)(1,0)$ for the conduction band calculation. Note that the density of states from the hole band merges into the density of states from the two conduction band ellipsoids centered on the k_z axis. The 1D density of states appears to be the appropriate quantity to consider since the wave function overlap in Eq. (12) strongly favors $\mathbf{k}$ corresponding to the conduction band minimum and valence band maximum. Localized states in the gap could assist the indirect Γ–Δ_4 tunneling, since a localized state is not restricted by the momentum conservation rules. This effect is not included in this model.

With our full-bandstructure model system, we now use the imaginary potential as a parameter to increase or decrease the magnitude of the band tails in the contacts and then observe the change in the shape and magnitude of the current as a function of applied voltage. A semiclassical Thomas–Fermi model is used to calculate the electrostatic potential.

III. NUMERICAL RESULTS AND DISCUSSION

Figure 4 displays I–V curves calculated for three values of electron lifetimes in the device contacts. The corresponding imaginary potentials are labeled on the figure. As the lifetime decreases, both the peak currents and valley currents increase. However, the valley current increases faster than the peak current resulting in a decrease in the peak-to-valley current ratio. As the lifetime decreases from 66 to 22 fs, the PVCR decreases from 7.3 to 2.7. There are two notable features shown in Fig. 4. The first is the qualitative agreement between the trends of the peak current, valley current, and post-valley turn-on current shown in this series of I–V and the series shown by Chynoweth *et al.*[15] which demonstrate the effect of increasing radiation dose on an Esaki diode. The increasing radiation dose alters the I–V by increasing the states in the band gap. The second important feature to note is that all of the current shown is temperature independent tunnel current. The turn-on of the current after the valley in
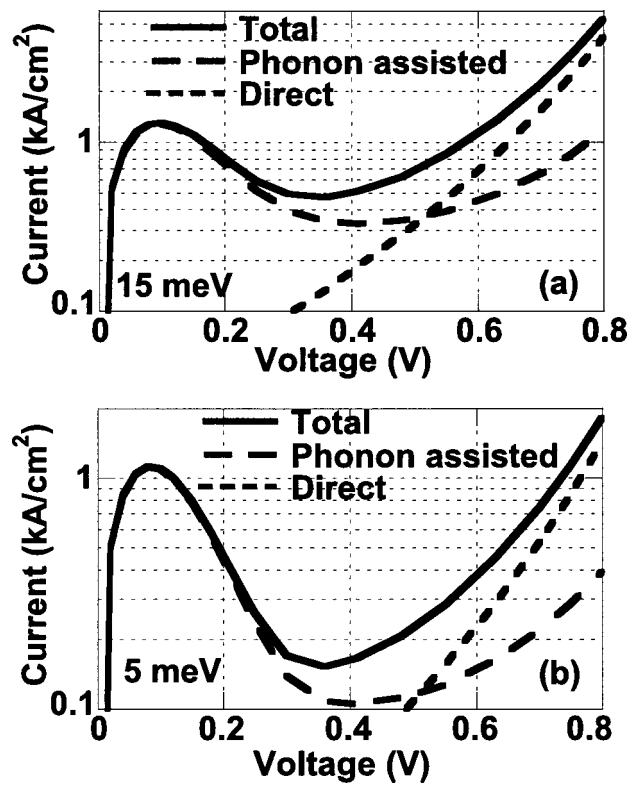

FIG. 5. Phonon-assisted and coherent components of the (a) 15 meV I–V and (b) 5 meV I–V from Fig. 4 plotted on a semilog scale.

Fig. 4 is still tunnel current. It is *not* the turn-on of the strongly temperature dependent diffusion current.

To explore further the physics of the peak current and excess current, we break the 5 and 15 meV I–V curves of Fig. 4 into their phonon-assisted and coherent components, and plot them in Fig. 5. The peak current is determined by the phonon-assisted tunneling current. However, the excess current, after the valley, is dominated by the direct tunneling current sourced from the bandtail states as sketched in Fig. 1.

Figure 6 displays the 700 °C anneal experimental I–V of Ref. 4 with the voltage axis representing the voltage across the diode corrected for the 12.2 Ω series resistance.[4] The voltage across the diode, V, is related to the applied voltage, V_a, by $V=V_a-12.2I$ where I is the current. The experimen-
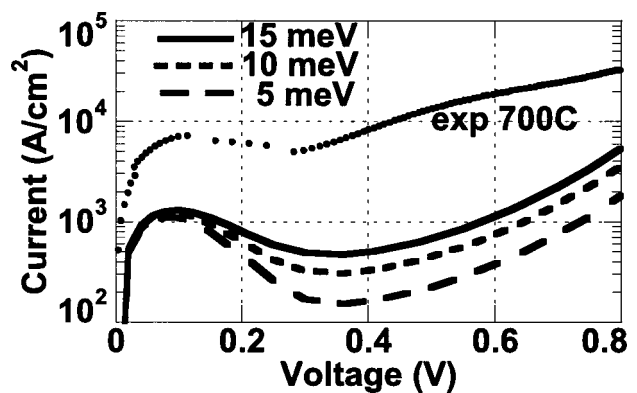

FIG. 6. Experimental 700 °C anneal I–V of Ref. 4, corrected for series resistance, overlaid on the calculated I–V of Fig. 4.

tal curve is overlaid on the I–V curves of Fig. 4. The theoretical I–V are calculated using the SIMS doping profile measured after the 700 °C anneal. There are similarities and differences between the calculated and experimental curves. The calculated voltage of both the peak current and valley current quantitatively matches the measured voltages (corrected for series resistance). The shoulder like feature in the negative differential resistance region of the experimental curves is the result of circuit oscillations during measurement and not a true representation of the intrinsic device current–voltage response.[40] The calculated current after the valley generally follows the trend of the measured current. However, all of the annealed experimental devices show a small hump in this region. This is referred to as the "hump" current, and it indicates the presence of a band of midgap states which are not included in our model.[41] The obvious discrepancy is the magnitude of the currents. The experimental peak current is 5.4 times larger than the calculated current.

There are several factors which affect the magnitude of the calculated current. The magnitude of the deformation potential has a quadratic effect on the current. The accuracy of the band model and the tunnel junction potential have an exponential effect on the current. First we consider the accuracy of the band model, particularly the evanescent dispersions at the conduction band minimum near X and the valence band maximum at Γ. We have numerically calculated the evanescent dispersions from our full band model and compared the results with those obtained from a single band model for the conduction and valence bands using the transverse electron mass of $0.19\,m_0$ and the light hole mass of $0.16\,m_0$, respectively. The results are shown in Fig. 7. As one would expect, the full band wave vectors are less than those calculated from the parabolic approximation. We compare the values at midgap, since that roughly corresponds to the middle of the tunnel barrier shown in Fig. 2. This is the point where the wave function envelope overlap calculated in Eq. (14) is maximum. At midgap, as shown in Fig. 7(a), the single-band conduction band evanescent wave vector differs from the full-band evanescent wave vector by 16%. At midgap, the single-band light-hole wave vector differs from the full-band light-hole wave vector by 30%. These trends are what one would qualitatively expect, however, we know of no method to experimentally, quantitatively verify the evanescent dispersion relations in the band gap for bulk Si.

One method to ascertain whether the decay through the gap is correct is to model the peak current versus intrinsic spacer layer length and compare to the experiment. However, the tunnel barrier which exponentially affects the current magnitude is determined by the dopant distributions through Poisson's equation. The dopant profiles vary on a nanometer scale which is difficult to resolve with state-of-the-art SIMS. The actual dopant profiles are sharper than the SIMS profiles. One test that we have made is to attempt to reproduce the peak–current versus intrinsic spacer–layer length shown in Fig. 3 of Ref. 6. Experimentally, for a series of four devices with nominal spacer layer lengths of 4, 6, 8, and 10 nm, all annealed at 650 °C for 1 min, the peak current followed an exponential dependence on the spacer layer length
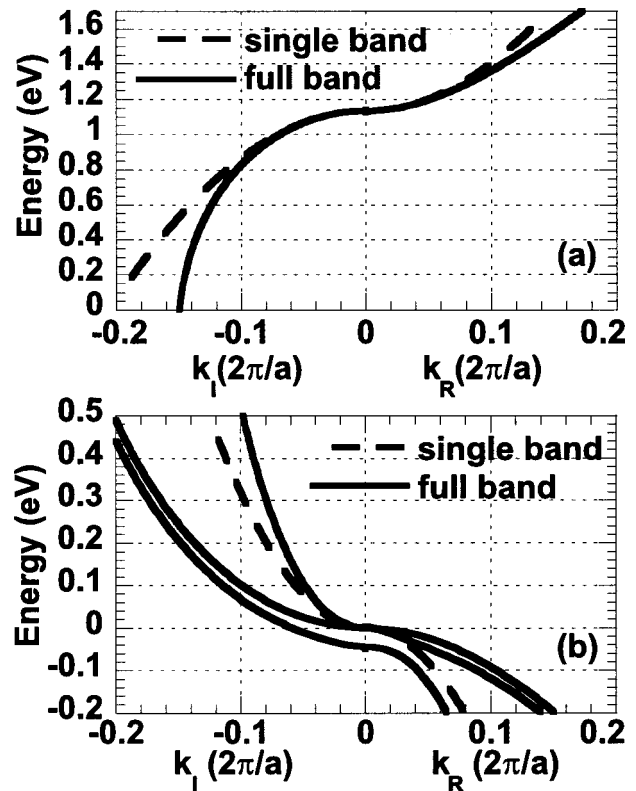

FIG. 7. Comparison of the real and imaginary dispersion relations calculated (a) from the parabolic single band model using the transverse conduction band mass of $0.19\,m_0$ and from the full-band model in the transverse mass direction of the Δ_4 conduction band valley and (b) from the single band model using the light hole mass of $0.16\,m_0$ and from the full-band model in the valence band in the (001) direction. The horizontal axis to the left of 0 is imaginary k and to the right of 0 is real k.

with a best fit dependence of $\exp(-2.0d)$ where d is the nominal spacer layer thickness in nanometers.

What we find numerically, using the SIMS doping profiles is very different. We start with the 700 °C doping profile from Ref. 4 that has been used for all of the I–V calculations presented so far. For that growth, the nominal distance between the Sb and B δ-doped regions is 6 nm. The Sb and B distribution are shown in Fig. 8(a) with the B distribution labeled by the nominal spacing of 6 nm. The actual peak-to-peak spacing shown is 9.1 nm. Since SIMS was not performed on the sequence of devices for which the experiments were performed in Ref. 6, we create a virtual experiment by shifting the known B SIMS distribution out in 2 nm increments. These are shown in Fig. 8(a) labeled by their nominal intrinsic layer distances of 8 and 10 nm. A closeup of the resulting tunnel junctions is shown in Fig. 8(b) where each conduction and valence band is labeled by the B SIMS profile in (a) that generated it. The tunnel distance t (nm) measured at the Fermi level is also shown for each junction. The tunnel distance increase is only 15% of the peak-to-peak doping distance. As the Sb–B spacing increases by 4 nm, the tunnel distance increases by only 0.6 nm. Thus, using these SIMS profiles, we would not expect to reproduce the $\exp(-2.0d)$ dependence of current versus nominal intrinsic layer distance shown in Fig. 3 of Ref. 6.

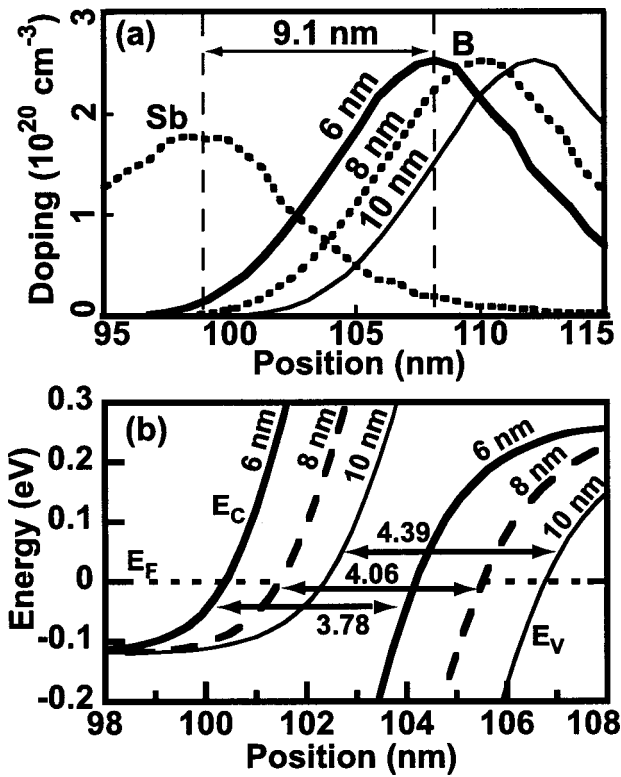A plot of the I–V corresponding to each doping profile is shown in Fig. 9. The peak current drops from 1300 to

FIG. 8. (a) Sb and B SIMS distributions for the 700 °C anneal from Ref. 4. The B distribution is labeled by the nominal distance between Sb and B δ doping of 6 nm. The actual peak-to-peak distance is 9.1 nm. This B distribution is then shifted to the right in 2 nm increments and labeled as 8 and 10 nm corresponding to our hypothetical experiment of increasing the nominal intrinsic layer distance. (b) Closeup of the tunnel junction resulting from the three B distributions shown in (a). The corresponding tunnel distance is shown.

FIG. 9. Three $I–V$'s corresponding to and labeled by the nominal intrinsic lengths of the three SIMS profiles and band diagrams shown in Fig. 8. The values of the peak current and PVCR for each curve are labeled.

310 A/cm^2 as the the B doping profile is shifted out by 4 nm. The PVCR initially is unchanged and then drops from 2.7 to 2.3. The reduction of the PVCR with tunnel distance has also been observed experimentally. If we plot the peak current versus nominal spacer distance, we obtain a dependence of $\exp(-0.36d)$. If we plot the peak current versus estimated tunnel distance shown in Fig. 8, we obtain a dependence of $\exp(-2.4d)$. We conclude from this that perhaps the largest error is introduced by the use of the raw SIMS data for quantitative comparisons with experiment.

In summary, the methods of modern quantum transport theory have been extended and applied to the problem of current transport in a LT-MBE grown Si tunnel diode. All three regions of the tunnel current are modeled: peak, valley, and post-valley turn-on. An increase in the magnitude of the band tails coincides with an increase in both the peak current and the valley current and a decrease in the PVCR. The presence of smooth band tails throughout the band gap in the contacts is sufficient to cause the turn-on of the tunnel current in the post-valley region of the $I–V$. This excess current is dominated by the direct coherent tunneling component of the current tunneling from gap state to gap state. The crossover between phonon assisted and direct occurs immediately after the valley minimum. The calculated voltages quantitatively match the experimental measurements. The magnitude of the calc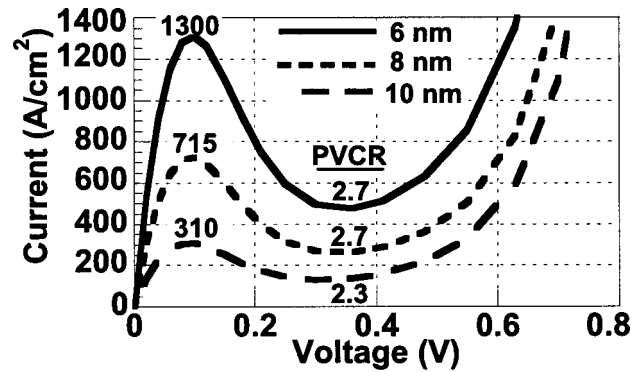ulated current is approximately a factor of 5.4 too small. Experimental and theoretical unknowns can account for this factor. We have considered sources of error arising from the dispersion relation of our model and the use of raw SIMS data, and we believe the the use of raw SIMS data gives rise to the largest source of error in these calculations.

ACKNOWLEDGMENTS

APPENDIX A: SURFACE GREEN FUNCTION CALCULATION

The calculation of the Γ in Eqs. (8)–(9) requires the calculation of the surface Green functions at sites 0 and $N+1$. This, in turn, requires calculating the surface Green functions of the semi-infinite, flat-band bulk material. These Green functions are then "walked in" using the recursive Green function algorithm to the edge of the nonequilibrium region shown in Fig. 2 and described in Ref. 29. Here, we describe the calculation of the surface Green functions of the semi-infinite bulk since the expressions that we use are different from those written down in Refs. 28 or 29. It is also an important topic since this calculation tends to take up a reasonable percentage of the total cpu time and we find that this calculation invariably presents a stumbling block when writing a new code with a complicated band structure model.

In the second neighbor sp^{3}s* model, the equation that we solve for the surface Green function using the generalized eigenfunction approach is for the semi-infinite region on the left[28]

$$g_s = [E - D + i\eta - t_{n,n-1}\chi_L z_L^{-1}\chi_L^{-1}]^{-1} \qquad (A1)$$

and for the semi-infinite region on the right

$$g_s = [E - D + i\eta - t_{n,n+1}\chi_R z_R \chi_R^{-1}]^{-1}, \qquad (A2)$$

where we have used block matrix notation.[29] Each symbol represents a 20×20 (sp^{3}s*) matrix with the elements consisting of the planar orbitals of the two-atom monolayer with the monolayer labeled by the index n; E is the scalar energy times the identity matrix; D is the intramonolayer n,n block; t couples adjacent monolayers. The imaginary potential $i\eta$ is added to the diagonal. The columns of $\chi_{L(R)}$ are the normalized eigenvectors returned from the LAPACK generalized eigenvalue routine $zgegv$ which are propagating to the left (right), and $z_{L(R)}$ is the diagonal matrix of complex propagation factors corresponding to the columns of $\chi_{L(R)}$. The generalized eigenvalue equations are described by Boykin.[33] We find that the χ that result from our Green function derivation are the eigenvectors propagating away from the device as opposed to the eigenvectors propagating towards the device as written in Eq. (B7) of Ref. 28, and that the corresponding propagation factors are z_L^{-1} on the left and z_R on the right. For example, on the left, the choice arises when one relates $g_{-1,0}$ to $g_{0,0}$. The physical interpretation of $g_{-1,0}$ is the system response at layer -1 to an excitation (particle injection) at layer 0. The causal response of the system is carried by the left propagating wave.

For the nearest neighbor sp^{3}s*d^5 model, it is numerically more efficient to work with atomic layers rather than monolayers. We label the two atomic layers in the basis as the a layer and the c layer in that order from left to right (the notation is brought over from the III–Vs where the two layers are the anions and the cations). The generalized eigenvalue equation is first order in z as opposed to second order in the second neighbor model.[33] It is obtained by writing down the equations of motion for the a and c layers of monolayer 0:

$$t_{0a,-1c}|-1,\chi_c\rangle+(D_a-E)|0,\chi_a\rangle+D_{0a,0c}|0,\chi_c\rangle=0, \tag{A3}$$

$$D_{0c,0a}|0,\chi_a\rangle+(D_c-E)|0,\chi_c\rangle+t_{0c,1a}|1,\chi_a\rangle=0. \tag{A4}$$

The subscript $0c$ indicates the c atomic layer of monolayer 0. Similarly, $|0,\chi_a\rangle$ is the a atomic layer component of χ in monolayer 0. Using the relation

$$|n+1,\chi\rangle=z|n,\chi\rangle \tag{A5}$$

in Eqs. (A3) and (A4) and combining them into a single matrix equation, we obtain the equation that we numerically solve

$$[\mathbf{A}-z\mathbf{B}]\begin{bmatrix}\chi_a\\\chi_c\end{bmatrix}=0, \tag{A6}$$

where

$$\mathbf{A}=\begin{bmatrix}D_{c,a} & (D_c-E)\\0 & t_{a,c}\end{bmatrix}, \tag{A7}$$

and

$$\mathbf{B}=-\begin{bmatrix}t_{c,a} & 0\\(D_a-E) & D_{a,c}\end{bmatrix}. \tag{A8}$$

One can also multiply Eq. (A6) through by $\mathbf{B}^{-1}$ and solve a standard eigenvalue equation. We have used both approaches.

We need the surface Green function $g_{c,c}$ for the last atomic layer of the semi-infinite region to the left and $g_{a,a}$ for the first atomic layer of the semi-infinite region to the right. Knowing χ from Eq. (A6), we then solve for the left

$$g_{c,c}=[E-D_c+i\eta-D_{c,a}(E-D_a)^{-1}(D_{a,c}+t_{a,c}\chi_{c,L}z_L^{-1}\chi_{c,L}^{-1})]^{-1}, \tag{A9}$$

and for the right

$$g_{a,a}=[E-D_a+i\eta-D_{a,c}(E-D_c)^{-1}(D_{c,a}+t_{c,a}\chi_{a,R}z_R\chi_{a,R}^{-1})]^{-1}. \tag{A10}$$

Equations (A9) and (A10) are derived in the same manner as Eqs. (A1), (A2), and (A6) starting from the equations of motion of the Green function. For example, for the right semi-infinite region, labeling the first monolayer as 0, we have

$$(E-D_a)g_{a,a}-D_{a,c}g_{c,a}=1, \tag{A11}$$

and

$$-D_{c,a}g_{a,a}+(E-D_c)g_{c,a}-t_{c,a1}g_{a1,a}=0. \tag{A12}$$

In Eqs. (A11) and (A12), the subscripts a and c indicate matrix elements of the a and c planar orbitals in monolayer 0. The last term in Eq. (A12) has a subscript $a1$ indicating matrix elements of the a layer of monolayer 1. We close the equations using

$$g_{a1,a}=\chi_{a,R}z_R\chi_{a,R}^{-1}g_{a,a}, \tag{A13}$$

use Eq. (A12) to write $g_{c,a}$ in terms of $g_{a,a}$, and substitute that expression into Eq. (A11) to obtain (A10). In Eqs. (A9)–(A13), all elements are 20×20 matrices corresponding to the ten planar atomic orbitals times 2 spins. D_c and D_a are the intra-atomic layer Hamiltonian matrices which would be diagonal except for the spin-orbit matrix elements. $D_{a,c}$ couples the a layer to the c layer to its right (intramonolayer). $t_{a,c}$ couples the a layer to the c layer to the left (intermonolayer). $\chi_{a,R}$ is the a orbital component of χ_R.

The procedure for sorting the eigenvectors returned from $zgegv$ into left propagating and right propagating groups can be, in practice, somewhat tricky, and we have not seen it described in the literature. The algorithm that we use (working with double complex precision), first looks at the magnitude of the eigenvalues z. If $|z|<1.0-10^{-8}$, the corresponding eigenvector is right propagating, i.e., it is exponentially decaying to the right. If $|z|>1.0+10^{-8}$, the corresponding eigenvector is left propagating, i.e., it is exponentially decaying to the left. Otherwise, the eigenvector is left (right) propagating if

$$v=-2.0\,\mathrm{Im}\{\chi^\dagger t_{c,a}\chi z\} \tag{A14}$$

is less than (greater than) 0.

Finally, we note that the generalized eigenvalue approach for calculating the surface Green function is not always robust, particularly for the second neighbor sp^{3}s* model.[33,35] We found that for the second neighbor sp^{3}s* model, the LAPACK generalized eigenvalue routine $zgegv$ would fail at a number of points in the transverse Δ valley. When this occurred, we resorted to the iteration method,

which we found to be completely robust, but more computationally expensive then the generalized eigenvalue approach. For high kinetic energy points with a small imaginary potential, $\eta = 2$ meV, hundreds of iterations could be required before convergence was obtained. The convergence criterion was to require the fractional change in the trace of g_s to be less than 10^{-6}. The equation for the iterative calculation of the surface Green function for the semi-infinite region on the left is

$$g_s = [E - D + i\eta - t_{n,n-1} g_s t_{n-1,n}]^{-1} \qquad \text{(A15)}$$

written in block matrix notation where each symbol in Eq. (A15) represents a 20×20 (sp^{3}s*) or 40×40 (sp^{3}s*d^5) matrix with the elements consisting of the planar orbitals of the two-atom monolayer with the monolayer labeled by the index n. E is the scalar energy times the identity matrix; D is the intramonolayer n,n block; the t couple adjacent monolayers; and the imaginary potential $i\eta$ is added to the diagonal.

[1] S. L. Rommel, T. E. Dillon, M. W. Dashiell, H. Feng, J. Kolodzey, P. R. Berger, P. E. Thompson, K. D. Hobart, R. Lake, A. Seabaugh, G. Klimeck, and D. K. Blanks, Appl. Phys. Lett. **73**, 2191 (1998).

[2] S. L. Rommel, T. E. Dillon, P. R. Berger, R. Lake, P. E. Thompson, K. D. Hobart, A. C. Seabaugh, and D. S. Simons, in *1998 IEDM Technical Digest* (IEEE, New York, 1998), p. 1035.

[3] S. L. Rommel, T. E. Dillon, P. R. Berger, P. E. Thompson, K. D. Hobart, R. Lake, and A. Seabaugh, IEEE Electron Device Lett. **20**, 329 (1999).

[4] P. E. Thompson, K. D. Hobart, M. E. Twigg, G. G. Jernigan, T. E. Dillon, S. L. Rommel, P. R. Berger, D. S. Simons, P. H. Chi, R. Lake, and A. Seabaugh, Appl. Phys. Lett. **75**, 1308 (1999).

[5] R. Duschl, O. G. Schmidt, G. Reitemann, E. Kasper, and K. Eberl, Electron. Lett. **35**, 1111 (1999).

[6] S. L. Rommel, N. Jin, T. E. Dillon, S. J. D. Giacomo, J. Banyai, B. M. Cord, C. D'Imperio, D. J. Hancock, N. Kirpalani, V. Emanuele, P. R. Berger, P. E. Thompson, K. D. Hobart, and R. Lake, in *2000 58th Annual Device Research Conference Digest* (IEEE, New York, 2000), p. 159.

[7] R. Duschl, O. G. Schmidt, and K. Eberl, Appl. Phys. Lett. **76**, 879 (2000).

[8] R. Duschl, O. G. Schmidt, and K. Eberl, Physica E **7**, 836 (2000).

[9] M. W. Dashiell, R. T. Troeger, S. L. Rommel, T. N. Adam, P. R. Berger, C. Guedj, J. Kolodzey, A. C. Seabaugh, and R. Lake, IEEE Trans. Electron Devices **46**, 1707 (2000).

[10] K. Eberl, R. Duschl, O. G. Schmidt, U. Denker, and R. Haug, J. Cryst. Growth **227-228**, 770 (2001).

[11] N. Jin, P. R. Berger, S. L. Rommel, P. E. Thompson, and K. D. Hobart, Electron. Lett. **37**, 1412 (2001).

[12] K. D. Hobart, P. E. Thompson, S. L. Rommel, T. E. Dillon, and P. R. Berger, J. Vac. Sci. Technol. B **19**, 290 (2001).

[13] P. E. Thompson, N. Jin, R. Yu, S-Y. Chung, A. T. Rice, P. R. Berger, R. Lake, H. Chi, and D. S. Simons, in The Third International SiGe(C) Epitaxy and Heterostructure Conference Digest, 2003.

[14] N. Jin, S-Y. Chung, A. T. Rice, P. R. Berger, R. Yu, P. E. Thompson, and R. Lake (unpublished).

[15] A. G. Chynoweth, W. L. Feldmann, and R. A. Logan, Phys. Rev. **121**, 684 (1961).

[16] L. V. Keldysh, Zh. Eksp. Teor. Fiz. **33**, 994 (1957); **34**, 962 (1958) [Sov. Phys. JETP **6**, 763 (1958); **7**, 665 (1958)].

[17] E. O. Kane, J. Phys. Chem. Solids **12**, 181 (1959); J. Appl. Phys. **32**, 83 (1961).

[18] P. J. Price and J. M. Radcliffe, IBM J. Res. Dev. **3**, 364 (1959).

[19] D. R. Fredkin and G. H. Wannier, Phys. Rev. **128**, 2054 (1962).

[20] E. Yang, J. Phys. Chem. Solids **36**, 1255 (1975).

[21] A. D. Carlo, P. Lugli, and P. Vogl, Solid State Commun. **101**, 921 (1997).

[22] W. H. Richardson, Appl. Phys. Lett. **73**, 1116 (1998).

[23] G. Reitemann and E. Kasper, Appl. Phys. Lett. **80**, 1294 (2002).

[24] T. B. Boykin, J. P. A. van der Wagt, and J. S. Harris, Phys. Rev. B **43**, 4777 (1991).

[25] D. Z. Y. Ting, E. T. Yu, and T. C. McGill, Phys. Rev. B **45**, 3583 (1992).

[26] J. A. Støvneng and P. Lipavsky, Phys. Rev. B **49**, 16494 (1994).

[27] A. D. Carlo, P. Vogl, and W. Potz, Phys. Rev. B **50**, 8358 (1994).

[28] R. C. Bowen, W. R. Frensley, G. Klimeck, and R. K. Lake, Phys. Rev. B **52**, 2754 (1995).

[29] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, J. Appl. Phys. **81**, 7845 (1997).

[30] C. Caroli, R. Combescot, P. Nozieres, and D. Saint-James, J. Phys. C **5**, 21 (1972).

[31] C. Rivas, R. Lake, G. Klimeck, W. R. Frensley, M. V. Fischetti, P. E. Thompson, S. L. Rommel, and P. R. Berger, Appl. Phys. Lett. **78**, 814 (2001).

[32] P. E. Thompson, K. D. Hobart, M. E. Twigg, S. L. Rommel, N. Jin, P. R. Berger, R. Lake, A. C. Seabaugh, P. H. Chi, and D. S. Simons, Thin Solid Films **380**, 145 (2000).

[33] T. B. Boykin, Phys. Rev. B **54**, 8107 (1996).

[34] J.-M. Jancu, R. Scholz, F. Beltram, and F. Bassani, Phys. Rev. B **57**, 6493 (1998).

[35] G. Klimeck, R. C. Bowen, T. B. Boykin, C. Salazar-Lazaro, T. A. Cwik, and A. Stoica, Superlattices Microstruct. **27**, 77 (2000).

[36] T. B. Boykin, Phys. Rev. B **60**, 15810 (1999).

[37] A. G. Chynoweth, R. A. Logan, and D. E. Thomas, Phys. Rev. **125**, 877 (1962).

[38] M. V. Fischetti and S. E. Laux, J. Appl. Phys. **80**, 2234 (1996).

[39] R. A. Logan and A. G. Chynoweth, Phys. Rev. **131**, 89 (1963).

[40] RCA Tunnel Diodes for Switching and Microwave Applications, Technical Manual TD-30, RCA Semiconductor and Materials Division, Somerville, NJ, 1963.

[41] S-Y. Chung, N. Jin, A. T. Rice, P. R. Berger, R. Yu, Z-Q. Fang, and P. E. Thompson, J. Appl. Phys. **93**, 9104 (2003).